Contents

1 Fundamentals

1.1. Set theory

1.1.1 Set brackets: { }. $A = \{a, b, c\}$ means that A is the set with elements a, b and c.

1.1.2 Set condition: is the vertical line |. $C = \{x \in A \,|\, (\text{condition on } x)\ \}$ indicates the subset C of A where all x satsify the given conditions.

1.1.3 Probability conditioning: $P(A|B)$ - the probability of A, *given that* B is the case. Notice that | in $P(A|B)$ is *not* a set operation. See (3).

1.1.4 Intersection may be written in two ways: $A \cap B$ or AB, and equals $\{x | x \in A \text{ and } x \in B\}$.

1.1.5 Union is written $A \cup B$, and equals $\{x | x \in A \text{ or } x \in B\}$.

1.1.6 Set difference is written either $A \setminus B$ or $A - B$, and equals $\{x | x \in A \text{ but } x \notin B\}$

1.1.7 The Universe: Ω is the symbol of "the entire universe of discourse", all possible relevant elements.

1.1.8 $\emptyset = \{\}$ is the empty set, that has no elements.

1.1.9 Complement: Given a universe Ω, then $A^c = \Omega \setminus A$. Reads "the complement of A".

1.1.10 Product: $A \times B = \{(a, b) | a \in A,\ B \in B\}$, the set of all ordered pairs (a, b).

1.1.11 $|A|$ or $n(A)$ means the cardinality of A; the number of its elements.

1.1.12 $|A \cup B| = |A| + |B| - |AB|$

1.1.13 $|AB| = |A| + |B| - |A \cup B|$

1.1.14 $|A - B| = |A| - |AB|$

1.1.15 $|A^c| = |\Omega| - |A|$

1.1.16 $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$ - the positive integers.

1.1.17 $\mathbb{N}_0 = \{0, 1, 2, 3, 4, \ldots\}$ - the non-negative integers.

1.1.18 $\mathbb{Z} = \{-4, -3, -2, -1, 0, 1, 2, 3, 4, \ldots\}$ - the integers, both positive and negative.

1.1.19 $\mathbb{R}$ - de real numbers; all the numbers on the number line.

1.1.20 Intervals (a, b): The brackets [and] mean that the respective endpoints are included, whereas the brackets $\langle$ and $\rangle$ means they're excluded. The brackets (and) may be used for an imprecise specifications of an interval when the endpoints are immaterial.

1.2. Indexing and repeat operations

1.2.1 Indexing: The simplest form of indexing is *enumeration*, like $x_1, x_2, x_3, \ldots$, and $A_1, A_2, A_3, \ldots$. Other kinds of indexing are by date, $(x_{2009.08.27})$, by place (A_{Boston}), or multiple indexes like $a_{2,3}$ or a_{23} for elements of matrices, or $a_{\text{Boston},\,2009.08.27}$ for both time and place.

1.2.2 $\displaystyle\bigcup_{k=m}^{n} A_k = A_m \cup A_{m+1} \cup \cdots \cup A_{n-1} \cup A_n$

1.2.3 $\displaystyle\bigcap_{k=m}^{n} A_k = A_m \cap A_{m+1} \cap \cdots \cap A_{n-1} \cap A_n$

1.2.4 $\displaystyle\sum_{k=m}^{n} a_k = a_m + a_{m+1} + \cdots + a_{n-1} + a_n$. If $n < m$, then $\displaystyle\sum_{k=m}^{n} a_k = 0$

1.2.5 $\displaystyle\prod_{k=m}^{n} a_k = a_m \cdot a_{m+1} \cdots a_{n-1} \cdot a_n$. If $n < m$, then $\displaystyle\prod_{k=m}^{n} a_k = 1$

1.2.6 $A \times B = \{(a,b) | a \in A,\, b \in B\}$

1.2.7 $\displaystyle\prod_{k=1}^{n} A_k = A_1 \times A_2 \times \cdots A_n = \{(a_1, a_2, \ldots, a_n) | a_k \in A_k\}$

1.2.8 $|A \times B| = |A| \cdot |B|$

1.2.9 $\displaystyle\left|\prod_{k=1}^{n} A_k)\right| = \prod_{k=1}^{n} |A_k|$

1.3. Frequently used formulas and functions

1.3.1 Factorial is defined for all non-negative integers $n \in \mathbb{N}_0$, and is: $\displaystyle n! = \prod_{k=1}^{n} k$

1.3.2 The Gamma function is factorial generalised. $\Gamma(n) = (n-1)!$ when $n \in \mathbb{N}$

1.3.3 Gamma for half values: $\Gamma(n + \frac{1}{2}) = \frac{(2n)!}{n!4^n}\sqrt{\pi}$ when $n \in \mathbb{N}$

1.3.4 Stirling's approximation I: $n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$

1.3.5 Stirling's approximation II: $n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \cdot \left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{1}{51840n^3} \cdots\right)$

1.3.6 Binomial: $\displaystyle\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$. The binomial is a special case of multinomial: $\binom{n}{k} = \binom{n}{k, n-k}$.

1.3.7 Multinomial: $\displaystyle\binom{n}{k_1, k_2, \ldots, k_m} = \frac{n!}{k_1! \cdot k_2! \cdot \cdots \cdot k_m!}$

1.3.8 Rewriting: $\displaystyle\binom{n}{k_1, k_2, \ldots, k_m} = \binom{n}{k_1} \cdot \binom{n-k_1}{k_2} \cdot \cdots \cdot \binom{n - k_1 - k_2 - \cdots - k_{m-2}}{k_{m-1}}$

1.3.9 TI/CASIO: n nCr k for $\binom{n}{k}$, og n nPr k for $\frac{n!}{(n-k)!}$ (HP: comb(n,k) og perm(n,k))

1.3.10 Pochhammer: $(a)_b = \frac{\Gamma(a+b)}{\Gamma(a)} = (b+a-1)$ nPr b (HP: perm$(b+a-1, b)$)

1.3.11 Pascal's triangle: $(a+b)^n = \sum_{k=1}^{n} \binom{n}{k} a^{n-k} b^k$

1.3.12 The incomplete Euler Beta function, for $x \in [0,1]$: $B_{(a,b)}(x) = \int_0^x t^{a-1}(1-t)^{b-1} dt$

1.3.13 The Euler Beta function: $B(a,b) = B_{(a,b)}(1) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

1.3.14 The Regularised Beta function, for $x \in [0,1]$: $I_{(a,b)}(x) = \frac{B_{(a,b)}(x)}{B_{(a,b)}} = \int_0^x \beta_{(a,b)}(t) dt$

1.3.15 The Beta function, for $t \in [0,1]$: $\beta_{(a,b)}(t) = \frac{1}{k} \cdot t^{a-1}(1-t)^{b-1}$ for $t \in [0,1]$; where $k = B(a,b)$

1.4. Sums

1.4.1 $\sum_{k=0}^{n} k = \frac{n(n+1)}{2}$

1.4.2 $\sum_{k=0}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$

1.4.3 $\sum_{k=0}^{n} k^3 = \frac{n^2(n+1)^2}{4}$

1.4.4 $\sum_{k=0}^{n} k^4 = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}$

1.4.5 $\sum_{k=0}^{n} k^5 = \frac{n^2(n+1)^2(2n^2+2n-1)}{12}$

1.4.6 Geometric series: $\sum_{k=A}^{B} r^k = \frac{r^A - r^{B+1}}{1-r}$

1.4.7 Infinite geometric series: $\sum_{k=A}^{\infty} r^k = \frac{r^A}{1-r}$ (iff $|r| < 1$)

1.4.8 $\sum_{k=0}^{\infty} kr^k = \frac{r}{(1-r)^2}$ (iff $|r| < 1$)

1.4.9 $\sum_{k=0}^{\infty} k^2 r^k = \frac{r(1+r)}{(1-r)^3}$ (iff $|r| < 1$)

1.4.10 $\sum_{k=1}^{\infty} \frac{r^k}{k} = -\ln(1-r)$ (iff $|r| < 1$)

1.4.11 $\sum_{k=0}^{\infty} \frac{r^k}{k!} = e^r$

2 Data: Measures of location and spread

- Individual data: $x_1, x_2, \dots, x_n$, or sorted: $x_{(1)}, x_{(2)}, \dots, x_{(n)}$
- Frequency data: Values $v_1, \dots, v_k$ with frequencies resp. $a_1, \dots, a_k$, totalt n data.
- Proportional data: Values $v_1, \dots, v_k$ with proportions resp. $p_1, \dots, p_k$
- Grouped data: The intervals go from l_k to $u_k = l_{k+1}$ and the number of measurements in interval k is a_k. Interval k has midpoint $v_k = \frac{l_k + u_k}{2}$ and width $b_k = u_k - l_k$. In total $n = \sum_k a_k$ measurements. Cumulative proportion: $A_k = p_1 + \cdots + p_k$ (Let $A_0 = 0$).

2.1. Proportional measures

	Individual data	Grouped data
Percentile P_p	Find $\kappa = \frac{p}{100} \cdot (n+1)$, and call the integer part h and the decimal part d. Then 2.1.1 $P_p = x_{(h)} + d \cdot (x_{(h+1)} - x_{(h)})$	Find k so that $A_{k-1} \leq \frac{p}{100} \leq A_k$. Then 2.1.2 $P_p = l_k + \frac{\frac{p}{100} - A_{k-1}}{p_k} \cdot (u_k - l_k)$
Median	2.1.3 $\tilde{x} = P_{50}$	
Quartiles	2.1.4 $Q_1 = P_{25}, Q_2 = P_{50}, Q_3 = P_{75}$	
Interquartile range	2.1.5 $Q_3 - Q_1$	
Median (simplified)	For individual data, the formula for median simplifies to: 2.1.6 $\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{if } n \text{ is odd} \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right) & \text{if } n \text{ is even} \end{cases}$	

2.2. Weighted measures

	Individual values	Frequencies and values	Proportions and values	Grouped values
Sum				
Σ_x	2.2.1 $\sum_{i=1}^{n} x_i$	2.2.2 $\sum_{j=1}^{k} a_j v_j$	2.2.3 $n \cdot \sum_{j=1}^{k} p_j v_j$	2.2.4 $\sum_{j=1}^{k} a_j v_j$
Sum of squares				
Σ_{x^2}	2.2.5 $\sum_{i=1}^{n} x_i^2$	2.2.6 $\sum_{j=1}^{n} a_j v_j^2$	2.2.7 $n \cdot \sum_{j=1}^{n} p_j v_j^2$	2.2.8 $\sum_k a_k \cdot (v_k^2 + \frac{1}{12} \cdot b_k^2)$

2.2.9 Mean: $\bar{x} = \frac{\Sigma_x}{n}$. For individual values we often use $\bar{x} = \frac{x_1 + \cdots + x_n}{n}$.

2.2.10 Total variation: $SS_x = \sum_k (x_k - \bar{x})^2 = \Sigma_{x^2} - n \cdot \bar{x}^2 = \Sigma_{x^2} - \frac{\Sigma_x^2}{n}$

2.2.11 Population variance: $\sigma_x^2 = \frac{SS_x}{n}$. For individual values we often use $\sigma_x^2 = \frac{1}{n} \sum_{k=1}^{n} (x_k - \bar{x})^2$.

2.2.12 Sample variance: $s_x^2 = \frac{SS_x}{n-1}$. For individual values we often use $s_x^2 = \frac{1}{n-1} \sum_{k=1}^{n} (x_k - \bar{x})^2$.

2.2.13 Population standard deviation: $\sigma_x = \sqrt{\sigma_x^2}$

2.2.14 Sample standard deviation: $s_x = \sqrt{s_x^2}$

2.3. Formulas for pairs $\{(x_i, y_i)\}_{i=1}^{n}$

2.3.1 Product sum: $\Sigma_{xy} = \sum_{k=1}^{n} x_k y_k$

2.3.2 Total joint variation: $SS_{xy} = \Sigma_{xy} - n \cdot \bar{x} \cdot \bar{y} = \Sigma_{xy} - \frac{\Sigma_x \cdot \Sigma_y}{n}$

2.3.3 Population covariance: $\sigma_{xy} = \frac{SS_{xy}}{n}$

2.3.4 Sample covariance: $s_{xy} = \frac{SS_{xy}}{n-1}$

2.3.5 Correlation: $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{SS_{xy}}{\sqrt{SS_x SS_y}} = \frac{s_{xy}}{s_x s_y} = r_{xy}$

3 Probability

Axioms	3.1.1(Kolmogorov's axioms, conditional version) $0 \leq P(A\|B) \leq 1$ $P(\Omega\|B) = 1$ $P(\bigcup_{i\in I} A_i\|B) =^1 \sum_{i\in I} P(A_i\|B)$ 3.1.2("Bayesian axioms") $P(AC\|B) = P(A\|B) \cdot P(C\|AB)$ $P(A\|B) + p(A^c\|B) = 1$ $P(B^c\|B) = 0$
Often used	3.1.3 $P(A^c) = 1 - P(A)$ 3.1.4 $P(A \cup B) = P(A) + P(B) - P(AB)$
Simplest model	3.1.5(Bayesians) $P(A) = P(A\|\Omega) = \frac{\text{favourable}}{\text{possible}} = \frac{n(A)}{n(\Omega)}$ 3.1.6(Frequentist) $P(A) = \lim_{n\to\infty} \frac{a(n)}{n}$ $a(n)$ = number of hits in A in n tries
Conditional	3.1.7 $P(A\|B) = \frac{P(AB)}{P(B)}$ 3.1.8 $P(AB) = P(A\|B)P(B)$ 3.1.9(Bayes' formula) $P(A\|B) = \frac{P(B\|A)P(A)}{P(B)}$
Independence	A and B are independent iff 3.1.10 $P(A\|B) = P(A)$ 3.1.11 $P(AB) = P(A)P(B)$
Conditional Independence	A and B are independent *given* C iff 3.1.12 $P(A\|BC) = P(A\|C)$ 3.1.13 $P(AB\|C) = P(A\|C)P(B\|C)$
Independence With many sets	A is independent of $B_1, \ldots, B_n$ iff 3.1.14 $P(A\|B_1B_2\cdots B_n) = P(A)$ and the equality also holds if you swap one or more of the B_k with B_k^c. $A_1, \ldots, A_n$ are independent, given B, iff 3.1.15 $P(A_1\cdots A_n\|B) = P(A_1\|B)\cdots P(A_n\|B)$ and the equality also holds if you swap one or more of the A_k with A_k^c.

[1] If all the A-s are disjoint and the index set I is a *countable*. Disjoint=mutually exclusive. This means that for $i \neq j$, $A_iA_j = \emptyset$. A set M is *countable* if it can be put into 1-1 correspondence with (a starting segment of) the integers, such that "1, 2, . . ." counts through the entire set. If the set is exhausted at some number, the set is *finite*. Otherwise, it is *countably infinite*.

4 Repeated sampling

4.0.1 Definition

- A **sequence** of k samples means the outcomes listed in the order in which they were sampled. It is *ordered.*
- A **combination** of k samples is a simple statement of how many were sampled of each kind. It is *unordered.*

4.0.2 Which kind of sampling is it?

To find which kind of sampling it is, you need to ask the following two questions:

- **With/without replacement:** Are the same elements available for the next sampling? This is what *with replacement* means. For dice and coins, this is the default. For sampling from an urn, this is the case only if you *replace* the elements before the next sampling.
  - Yes: "with replacement"
  - No: "without replacement"
- **Ordered/Unordered:** Does the order of the results matter?
  - Yes: "ordered" – a sequence, like Head-Tails-Tails-Head-Head
  - No: "unordered" – a combination, like 3 Heads, 2 Tails

4.1. Combinatorial:

Number possible samplings

To find the parameters n and k for the formula:

- How many elements are available on the first sampling? This is n.
- How many tries? This is k.

	With replacement	Without replacement
Ordered	4.1.1 n^k	4.1.2 $\frac{n!}{(n-k)!}$
Unordered	4.1.3 $\binom{n+k-1}{k}$	4.1.4 $\binom{n}{k}$

4.2. Probabilistic: Sampling from a population of 2 kinds;

- N = total number of elements in population (if finite).
- $S = S_1$ = total number of elements of the first kind in the population.
- $p = p_1$ $\left(= \frac{S_1}{N}\right)$ = proportion of elements of the first kind in the population.
- n = number of trials.
- $k = k_1$ = number of results from the first kind.

The probability of such a result is then:

	With replacement	**Without** replacement
Sequence (ordered)	4.2.1 $p^k(1-p)^{n-k}$	4.2.2 $\frac{\binom{N-n}{S-k}}{\binom{N}{S}}$
Combination (unordered)	4.2.3 $\binom{n}{k}p^k(1-p)^{n-k}$	4.2.4 $\frac{\binom{S}{k}\binom{N-S}{n-k}}{\binom{N}{n}} = \frac{\binom{N-n}{S-k}\binom{n}{k}}{\binom{N}{S}}$

When N is large, the formulas for sampling *with* replacement give good approximations for sampling *without* replacement.

4.3. Probabilistic: Sampling from a population of n kinds;

- N = total number of elements in population (if finite).
- S_j = total number of elements of the j-th kind in the population.
- p_j $\left(= \frac{S_j}{N}\right)$ = proportion of elements of the j-th kind in the population.
- n = number of trials.
- k_j = number of results from the j-th kind.

The probability of such a result is then:

	With replacement	**Without** replacement
Sequence (ordered)	4.3.1 $p_1^{k_1}p_2^{k_2}\cdots p_m^{k_m}$	4.3.2 $\frac{\binom{N-n}{S_1-k_1,S_2-k_2,\ldots,S_m-k_m}}{\binom{N}{S_1,S_2,\ldots,S_m}}$
Combination (unordered)	4.3.3 $\binom{n}{k_1,k_2,\ldots,k_m}p_1^{k_1}p_2^{k_2}\cdots p_m^{k_m}$	4.3.4 $\binom{n}{k_1,k_2,\ldots,k_m}\cdot\frac{\binom{N-n}{S_1-k_1,S_2-k_2,\ldots,S_m-k_m}}{\binom{N}{S_1,S_2,\ldots,S_m}}$

When N is large, the formulas for sampling *with* replacement give good approximations for sampling *without* replacement.

5 Stochastic variables

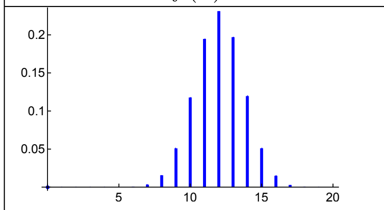
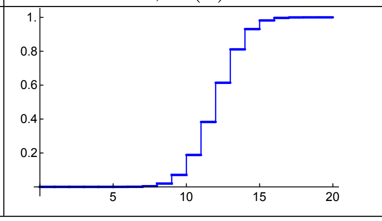
	Discrete distribution, $f(x) = p_x$	Continuous distribution $f(x)$
Probability, method 1	5.1.1 $P(X \in A) = \sum_{x \in A} p_x$	5.1.2 $P(X \in (a,b)\,) = \int_a^b f(t)dt$
Cumulative distribution	5.1.3 $F(x) = \sum_{t \leq x} p_t$	5.1.4 $F(x) = \int_{-\infty}^{x} f(t)dt$
1. moment, Expected value $\mu_X = E[X]$	5.1.5 $E[X] = \mu_X = \sum_{\text{all } x} xp_x$	5.1.6 $E[X] = \mu_X = \int_{-\infty}^{\infty} x \cdot f(x)dx$
2. moment, $E[X^2]$	5.1.7 $E[X^2] = \sum_{\text{all } x} x^2 p_x$	5.1.8 $E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx$
General $E[h(X)]$	5.1.9 $E[h(X)] = \sum_{\text{alle } x} h(x) \cdot p_x$	5.1.10 $E[h(X)] = \int_{-\infty}^{\infty} h(x) \cdot f(x)dx$
$E[XY]$	5.1.11 $E[XY] = \sum_{\text{alle } x,y} xy \cdot p_{xy}$	5.1.12 $E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f(x,y)dydx$
Probability, method 2	5.1.13 $P(X \in \langle a,b]) = F(b) - F(a)$	
Variance σ^2	5.1.14 $\sigma_X^2 = Var(X) = E[X^2] - \mu_X^2$	
Standard deviation σ **and precision** τ	5.1.15 $\sigma_X = \sqrt{\sigma_X^2}$	5.1.16 $\tau_X = \frac{1}{\sigma_X^2}$
Covariance σ **and correlation** ρ	5.1.17 $\sigma_{XY} = Cov(X,Y) = E[XY] - \mu_X \mu_Y$	5.1.18 $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$
Percentile and Median	5.1.19 P_p is the x that solves $F(x) = \frac{p}{100}$	5.1.20 The median is $\tilde{x} = P_{50}$

6 Probability distributions

6.1. Discrete probability distributions

Probability distributions on CASIO: OPTN » STAT » DIST
Probability distributions on Texas Instruments: 2nd » VARS (DIST)

6.1.1 HYPERGEOMETRIC DISTRIBUTION, $\text{hyp}_{(n,S,N)}(x)$, $x \in \{0, 1, \ldots, n\}$

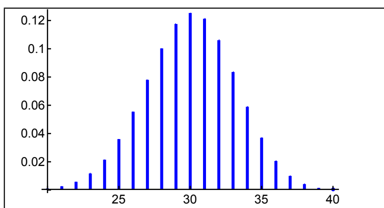
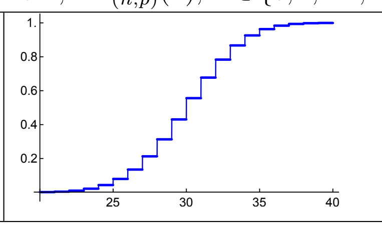
Distribution $f(x)$	Cumulative, $F(x)$	Formulas
		pdf: $X \sim \text{hyp}_{(n,S,N)}(x) = \frac{\binom{S}{x}\binom{N-S}{n-x}}{\binom{N}{n}}$ CDF: $\text{HYP}_{(n,S,N)}(x) = \sum_{z=0}^{x} \text{hyp}_{(n,S,N)}(z)$ $\mu_X = np$ der $p = \frac{S}{N}$ $\sigma_X^2 = np(1-p) \cdot \frac{N-n}{N-1}$

Mathematica: HypergeometricDistribution[n, S, N]

CASIO: H-GEO » $\text{hyp}_{(n,S,N)}(x) =$ Hpd →HypergeoPD(x, n, S, N)
$\text{HYP}_{(n,S,N)}(x) =$ Hcd →HypergeoCD(x, n, S, N)

Approximations: $\text{bin}_{(n,p)}(x)$ when $n < \frac{N}{10}$
$\text{pois}_{np}(x)$ when $100 < n < \frac{N}{10}$ and $n^{0.31}p < 0.47$
normal approximation (7.2) when $n < \frac{N}{10}$ and $np(1-p) > 5$.

6.1.2 BINOMIAL DISTRIBUTION, $\text{bin}_{(n,p)}(x)$, $x \in \{0, 1, \ldots, n\}$

		pdf: $X \sim \text{bin}_{(n,p)}(x) = \binom{n}{x} p^x (1-p)^{n-x}$ CDF: $\text{BIN}_{(n,p)}(x)$ $\mu_X = np$ $\sigma_X^2 = np(1-p)$

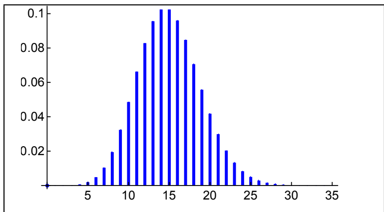
Special instance: When $n = 1$, we have the *Bernoulli distribution*: $\text{bern}_p = \text{bin}_{(1,p)}$.

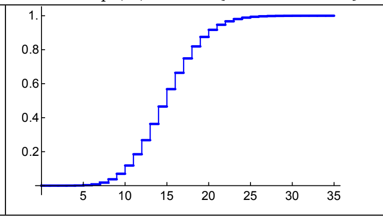
Mathematica: BinomialDistribution[n, p]

$\text{bin}_{(n,p)}(x) =$ HP: binomial(n, x, p)
CASIO: BINM » Bpd→ BinomialPD(x, n, p)
TI: binompdf→ binompdf(n, p, x)
$\text{BIN}_{(n,p)}(x) =$ HP: binomial_cdf(n, p, x)
CASIO: BINM » Bcd→ BinomialCD(x, n, p)
TI: binomcdf→ binomcdf(n, p, x)

Approximations: $\text{bin}_{(n,p)}(x) \approx \text{pois}_{np}(x)$ when $n^{0.31}p < 0.47$.
normal approximation (7.2) when $np(1-p) > 5$.

6.1.3 POISSON DISTRIBUTION, $\text{pois}_p(x)$, $x \in \{0, 1, \ldots, \infty\}$





pdf: $X \sim \text{pois}_\lambda(x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$

CDF: $\text{POIS}_\lambda(x) = \sum_{z=0}^{x} \text{pois}_\lambda(z)$

$\mu_X = \lambda$

$\sigma_X^2 = \lambda$

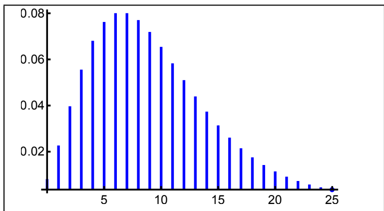
Mathematica: PoissonDistribution[λ]

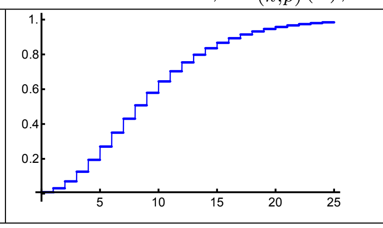
$\text{pois}_\lambda(x) =$ HP: poisson(λ, x)
CASIO: POISN » Ppd→ PoissonPD(x, λ)
TI: poissonpdf→ poissonpdf(λ, x)

$\text{POIS}_\lambda(x) =$ HP: poisson_cdf(λ, x)
CASIO: POISN » Pcd→ PoissonCD(x, λ)
TI: poissoncdf→ poissoncdf(λ, x)

Approximations: normal approximation (7.2) when $\lambda > 10$

6.1.4 NEGATIVE BINOMIAL DISTRIBUTION, $\text{nb}_{(k,p)}(x)$, $x \in \{0, 1, \ldots, \infty\}$





pdf: $X \sim \text{nb}_{(k,p)}(x) = \binom{x+k-1}{k-1} p^k (1-p)^x$

CDF: $\text{NB}_{(k,p)}(x) = \sum_{z=0}^{x} \text{nb}_{(k,p)}(z)$

$\mu_X = \frac{k(1-p)}{p}$

$\sigma_X^2 = \frac{k(1-p)}{p^2}$

Special instance: When $k = 1$, we have the *Geometric distribution* (geom/GEOM).

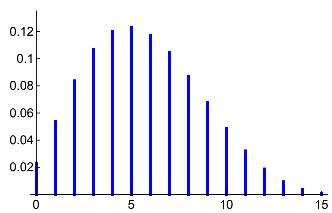
Mathematica: NegativeBinomialDistribution[k, p]

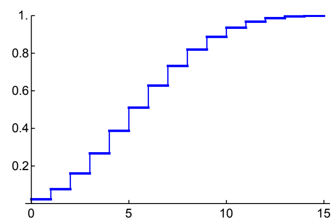
$\text{nb}_{(k,p)}(x) = p \cdot \text{bin}_{(x+k-1,p)}(k-1)$
= HP: $p \cdot \text{binomial}(x+k-1, k-1, p)$
CASIO: $p \cdot \text{BinomialPD}(k-1, x+k-1, p)$
TI: $p \cdot \text{binomial}(x+k-1, p, k-1)$

$\text{NB}_{(k,p)}(x) = 1 - \text{BIN}_{(x+k,p)}(k-1)$ when $x, k \in \mathbb{N}_0$
= HP: $1 - \text{binomial\_cdf}(x+k, p, k-1)$
CASIO: BINM » Bcd→ $1 - \text{BinomialCD}(k-1, x+k, p)$
TI: binomcdf→ $1 - \text{binomcdf}(x+k, p, k-1)$

$\text{NB}_{(k,p)}(x) = I_{(k,x+1)}(p)$ for all $x, k > 0$
HP: $\text{fisher\_cdf}(2k, 2(x+1), \frac{(x+1)p}{k(1-p)})$
CASIO: FCd→ $\text{FCD}(0, \frac{(x+1)p}{k(1-p)}, 2k, 2(x+1))$
TI: 0: → $\text{Fcdf}(0, \frac{(x+1)p}{k(1-p)}, 2k, 2(x+1))$

6.1.5 BETA BINOMIAL DISTRIBUTION, $\beta \mathrm{b}_{(a,b,n)}$





When $X \sim bin_{(n,p)}$ and $p \sim \beta_{(a,b)}$,
then $X \sim \beta b_{(a,b,n)}$

$$\text{pdf:} \quad X \sim \quad \beta b_{(a,b,n)}(x) = \binom{n}{x} \frac{(a)_x (b)_{n-x}}{(a+b)_n} \quad *$$

$$= \binom{n}{x} \cdot \frac{\binom{a+b}{a} \cdot \frac{ab}{a+b}}{\binom{a+b+n}{a+x} \cdot \frac{(a+x)(b+n-x)}{a+b+n}}$$

CDF: ß$\mathrm{B}_{(a,b,n)}(x) = \sum_{z=0}^{x} \beta b_{(a,b,n)}(z)$

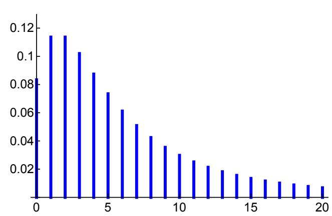
$$\mu_X = \frac{na}{a+b}$$

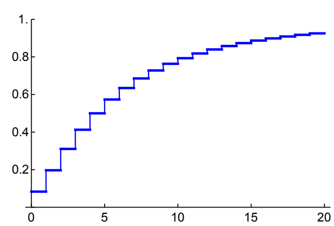
$$\sigma_X^2 = \frac{nab(a+b+n)}{(a+b)^2(a+b+1)}$$

$*$ Pochhammer: (1.3.10)

Mathematica: BetaBinomialDistribution[a,b,n]
Calculator: Use the formula above with (sums and) binomials.

6.1.6 BETA NEGATIVE BINOMIAL DISTRIBUTION, $\beta \mathrm{nb}_{(a,b,k)}$





When $X \sim nb_{(k,p)}$ and $p \sim \beta_{(a,b)}$,
then $X \sim \beta nb_{(a,b,k)}$

$$\text{pdf:} \quad X \sim \quad \beta nb_{(a,b,k)}(x) = \binom{x-1}{k-1} \frac{(a)_k (b)_x}{(a+b)_{x+k}} \quad *$$

$$= \binom{x+k-1}{k-1} \cdot \frac{\binom{a+b}{a} \cdot \frac{a+b+k+x}{a+b}}{\binom{a+b+k+x}{a+k} \cdot \frac{(a+k)(b+x)}{ab}}$$

CDF: ß$\mathrm{NB}_{(a,b,k)}(x) = \sum_{z=0}^{x} \beta nb_{(a,b,k)}(z)$

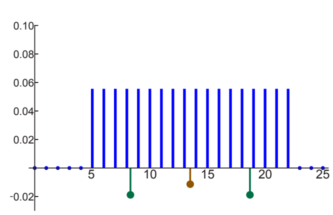
$$\mu_X = \frac{bk}{a-1}$$

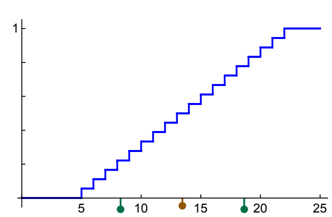
$$\sigma_X^2 = \frac{kb(a+b-1)(a+k-1)}{(a-2)(a-1)^2}$$

$*$ Pochhammer: (1.3.10)

Mathematica: BetaNegativeBinomialDistribution[a,b,k]
Calculator: Use the formula above with (sums and) binomials.

6.1.7 UNIFORM DISTRIBUTION, nb_p





Integers $x \in \{a, a+1, \ldots, b\}$; $n = b - a + 1$

pdf: $X \sim u_{(a,b)}(x) = \frac{1}{n}$

CDF: $U_{(a,b)}(x) = \frac{x-a+1}{n}$

$\mu_X = E[X] = \frac{a+b}{2}$

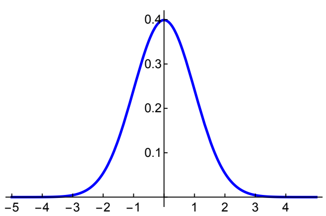
$\sigma_X^2 = \frac{n^2-1}{12}$

Mathematica: DiscreteUniformDistribution[min,max]

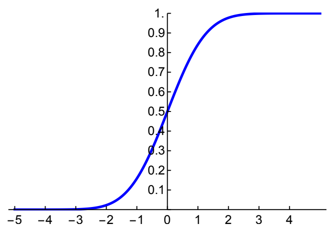
7 Continuous probability distributions

7.1. The Normal distribution

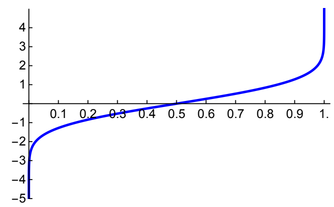
7.1.1 Probability density (pdf): $X \sim \phi_{(\mu,\sigma)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$



$\phi_{(0,1)}(x)$



$\Phi_{(0,1)}(x)$



$\Phi^{-1}_{(0,1)}(p) = z_p$

7.1.2 Expected value=median=mode: $\mu_X = E[X] = \widetilde{X} = X_{max} = \mu$

7.1.3 Variance: $\sigma_X^2 = \sigma^2$

7.1.4 Cumulative probability (CDF): $P(X \leq x) = \Phi_{(\mu,\sigma)}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

Mathematica:		CDF[NormalDistribution[μ, σ], x]
HP:		normald_cdf(μ, σ, x)
CASIO:	NORM » Ncd→	NormCD($-10^{99}, x, \sigma, \mu$)
TI:	normalcdf→	normalcdf($-10^{99}, x, \mu, \sigma$)

7.1.5 Inverse cumulative (INV): $\Phi^{-1}_{(\mu,\sigma)}(p) = \mu + z_p \cdot \sigma$ with shorthand $z_p = \Phi^{-1}_{(0,1)}(p)$

Mathematica:		InverseCDF[NormalDistribution[μ, σ], p]
HP:		normald_icdf(μ, σ, p)
CASIO:	NORM » InvN→	InvNormCD($-1, p, \sigma, \mu$)
TI:	invNorm→	invNorm(p, μ, σ)

7.2. Normal approximation

7.2.1 For continuous $X \sim f_X(x)$the normal approximation is:

$$f_X(a) \approx \phi_{(\mu_X,\sigma_X)}(a)$$

$$F_X(a) = P(X \leq a) \approx \Phi_{(\mu_X,\sigma_X)}(a)$$

7.2.2 For discrete $X \sim f_X(x)$the normal approximation is:

$$f_X(a) \approx \Phi_{(\mu_X,\sigma_X)}(a + \tfrac{1}{2}) - \Phi_{(\mu_X,\sigma_X)}(a - \tfrac{1}{2})$$

$$F_X(a) = P(X \leq a) \approx \Phi_{(\mu_X,\sigma_X)}(a + \tfrac{1}{2})$$

7.3. The sum of normal distributed stochastic variables

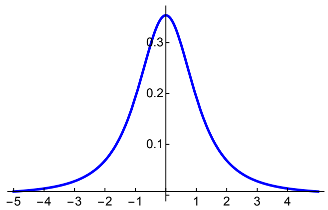
$X = a_1X_1 + \cdots + a_nX_n$, where a_k can take on any value, both positive and negative. Since the normal distribution is fully specified when you know μ and σ, then $X \sim \phi_{(\mu_X,\sigma_X)}$, where

7.3.1 $\mu_X = a_1\mu_{X_1} + \cdots + a_n\mu_{X_n}$

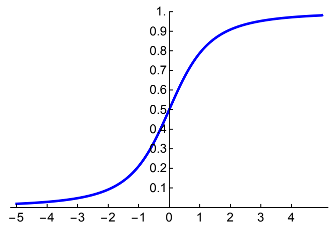
7.3.2 $\sigma_X^2 = a_1^2\sigma_{X_1}^2 + \cdots + a_n^2\sigma_{X_n}^2$

7.4. Student's t distribution

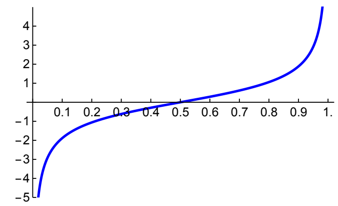
7.4.1 Probability density (pdf): $X \sim t_{(\mu,\sigma,\nu)}(x) = \left(\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \cdot \frac{1}{\sigma\sqrt{\pi\nu}}\right) \cdot \left(1 + \frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}$



$t_{(0,1)}(x)$



$T_{(0,1)}(x)$



$T^{-1}_{(0,1,\nu)}(p) = t_{\nu,p}$

7.4.2 Expected value=median=mode: $\mu_X = E[X] = \widetilde{X} = X_{max} = \mu$

7.4.3 Variance: $\sigma_X^2 = \begin{cases} \sigma^2 \frac{\nu}{\nu-2} & \nu > 2 \\ \infty & \nu \leq 2 \end{cases}$

7.4.4 Cumulative probability (CDF): $P(X \leq x) = T_{(\mu,\sigma,\nu)}(x) = T_\nu\left(\frac{x-\mu}{\sigma}\right)$

Mathematica:		CDF[StudentTDistribution[μ, σ, ν], x]
HP:		student_cdf(ν, $\frac{x-\mu}{\sigma}$)
CASIO:	t » Tcd→	tCD(-10^{99}, $\frac{x-\mu}{\sigma}$, ν)
TI:	tcdf→	tcdf(-10^{99}, $\frac{x-\mu}{\sigma}$, ν)

7.4.5 Inverse cumulative (INV): $T^{-1}_{(\mu,\sigma,\nu)}(p) = \mu + \sigma \cdot t_{\nu,p}$ with shorthand $t_{\nu,p} = T^{-1}_{(0,1,\nu)}(p)$

Mathematica:		InverseCDF[StudentTDistribution[μ, σ, ν], p]
HP:		$\mu + \sigma *$ student_icdf(ν, p)
CASIO:	t » InvT→	$\mu - \sigma *$ InvTCD(p, ν)
TI:	invT→	$\mu + \sigma *$ invT(p, ν)

7.5. Sum and difference of two t distributions: $Z = X \pm Y$

(Satterthwaite's approximation)

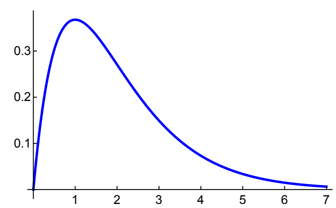
7.5.1 $\mu_Z = \mu_X \pm \mu_Y$

7.5.2 $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$

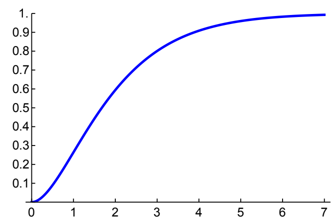
7.5.3 $\nu_Z = \left\lfloor \frac{\left(\frac{\sigma_X^2}{\nu_X+1} + \frac{\sigma_Y^2}{\nu_Y+1}\right)^2}{\left(\frac{\left(\frac{\sigma_X^2}{\nu_X+1}\right)^2}{\nu_X} + \frac{\left(\frac{\sigma_Y^2}{\nu_Y+1}\right)^2}{\nu_Y}\right)} \right\rfloor$ (where $\lfloor x \rfloor$ is the largest integer below or equal to x)

7.6. The gamma distribution

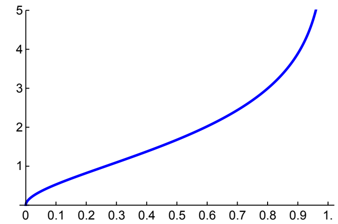
7.6.1 Probability density (pdf): $T \sim \gamma_{(k,\lambda)}(t) = \frac{(\lambda t)^{k-1}}{(k-1)!}\lambda \cdot e^{-\lambda t}$ for $t \in (0, \infty)$



$\gamma_{(k,\lambda)}(t)$



$\Gamma_{(k,\lambda)}(t) = \mathbb{X}^2_{(2k)}(2\lambda t)$



$\Gamma^{-1}_{(k,\lambda)}(p) = \frac{1}{2\lambda} \cdot \chi^2_{2k}(p)$

7.6.1 Expected value: $\mu_T = E[T] = \frac{k}{\lambda}$

7.6.2 Median: $\widetilde{T} = \Gamma^{-1}_{(\kappa\,,\,\tau)}(0.5)$

7.6.3 Mode: $T_{max} = \frac{\kappa-1}{\tau}$

7.6.4 Variance: $\sigma^2_T = \frac{k}{\lambda^2}$

7.6.2 Cumulative probability (CDF): $P(T \leq t) = \Gamma_{(k,\lambda)}(t) = 1 - \sum_{n=0}^{k-1} \frac{(\lambda t)^n}{n!} e^{-\lambda t}$ (when $n \in \mathbb{N}$)

$\Gamma_{(k,\lambda)}(t) =$	$(k > 0)$	Mathematica:		CDF[GammaDistribution[k, $\frac{1}{\lambda}$], t]
	$(k \in \mathbb{N})$	Mathematica:		CDF[ErlangDistribution[k, λ], t]
	$(2k \in \mathbb{N})$	HP:		chisquare_cdf($2k, 2\lambda t$)
	$(2k \in \mathbb{N})$	CASIO:	CHI » CCd→	ChiCD($0, 2\lambda t, 2k$)
	$(2k \in \mathbb{N})$	TI:	χ^2cdf→	χ^2cdf($0, 2\lambda t, 2k$)

7.6.3 Inverse cumulative (INV): $\Gamma^{-1}_{(k,\lambda)}(p)$ with shorthand $\frac{1}{2\lambda} \cdot \chi^2_{2k}(p)$

$\Gamma^{-1}_{(k,\lambda)}(p) =$	$(k > 0)$	Mathematica:		InverseCDF[GammaDistribution[k, $\frac{1}{\lambda}$], p]
	$(k \in \mathbb{N})$	Mathematica:		InverseCDF[ErlangDistribution[k, λ], p]
	$(2k \in \mathbb{N})$	HP:		chisquare_icdf($2k, p$)/(2λ)
	$(2k \in \mathbb{N})$	CASIO:	CHI » InvC→	InvChiCD($1 - p, 2k$)/(2λ)
	$(2k \in \mathbb{N})$	TI (CX):	Invχ^2 →	Invχ^2($p, 2k$)/(2λ)
	$(2k \in \mathbb{N})$	TI (83/84):	MATH » Solver:	χ^2cdf($0, 2\lambda x, 2k$) $- p = 0$ » $x = 1$ » Alpha » Enter

Special instances: When $n \in \mathbb{N}$, the gamma distribution is called a *Erlang distribution*.
When $n = 1$ the gamma distribution is called an *exponential distribution*.
When $2k \in \mathbb{N}$, then $\gamma_{(k,\frac{1}{2})}$ is the chi square (χ^2) distribution with $\nu = df = 2k$ degrees of freedom.

Approximations: normal approximation (7.2) when $k > 30$.

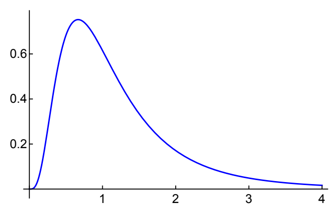
Relations:

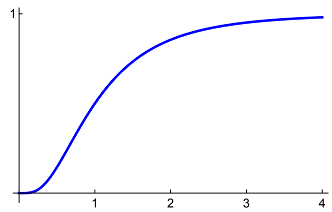
1. When $T \sim \gamma_{(k,\lambda)}$, and $\lambda \sim \gamma_{(\kappa,\tau)}$, then $T \sim g\gamma_{(k,\kappa,\tau)}(t)$.
2. If $T_1 \sim \gamma_{(k,\lambda)}(t)$, $T_2 \sim \gamma_{(\kappa,\tau)}(t)$, $m = \frac{\kappa\lambda}{k\tau}$, $Q = \frac{T_1}{T_2}$, and $Q_* = mQ$, then

$$Q_* \sim f_{(2k,2\kappa)}(t) \qquad Q \sim m \cdot f_{(2k,2\kappa)}(mt)$$
$$P(Q_* < t) = F_{(2k,2\kappa)}(t) \qquad P(Q < t) = F_{(2k,2\kappa)}(mt)$$

7.7. Other continuous probability distributions

7.7.1 F DISTRIBUTION, $f_{(\alpha,\beta)}(t)$, $t \in (0, \infty)$





pdf: $T \sim f_{(\alpha,\beta)}(t) = \dfrac{\sqrt{\frac{(\alpha t)^{\alpha}\cdot\beta^{\beta}}{(\alpha t+\beta)^{\alpha+\beta}}}}{t\cdot \mathrm{B}_{(\alpha/2\,,\,\beta/2)}}$

CDF: $F_{(\alpha,\beta)}(t) = \mathrm{I}_{(\alpha/2\,,\,\beta/2)}\left(\frac{\alpha t}{\alpha t+\beta}\right)$

$\mu_T = E[T] = \frac{\beta}{\beta-2}$

$\sigma_T^2 = \frac{2\beta^2(\alpha+\beta-2)}{\alpha(\beta-4)(\beta-2)^2}$

$\widetilde{T} = F^{-1}_{(\alpha,\beta)}(0.5)$

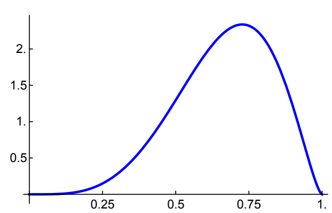
$T_{max} = \frac{\beta(\alpha-2)}{\alpha(\beta+2)}$

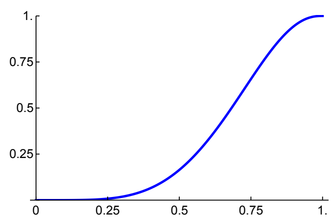
Mathematica: FRatioDistribution[α, β]

$F_{(\alpha,\beta)}(t) =$ HP: fisher_cdf(α, β, t)
CASIO: F » FCd→FCD($0, t, \alpha, \beta$)
TI: Fcdf→ Fcdf($0, t, \alpha, \beta$)

$F^{-1}_{(\alpha,\beta)}(p) =$ HP: fisher_icdf(α, β, p)
CASIO: F » InvF→ InvFCD($1-p, \alpha, \beta$)
TI (CX): InvF→ InvF(p, α, β)
TI (83/84): MATH » Solver: Fcdf($0, x, \alpha, \beta$) $- p = 0$ » ⇓ » $X = 0.5$ » Alpha » Enter

7.7.2 BETA DISTRIBUTION, $\beta_{(a,b)}(x)$, $x \in (0, 1)$





pdf: $X \sim \beta_{(a,b)}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$

CDF: $\mathrm{I}_{(a,b)}(x) = F_{(2a,2b)}\left(\frac{bx}{a(1-x)}\right)$

$\mu_X = \frac{a}{a+b}$

$\sigma_X^2 = \frac{ab}{(a+b)^2(a+b+1)}$

$\widetilde{X} = I^{-1}_{(a,b)}(0.5)$

$X_{max} = \frac{a-1}{a+b-2}$

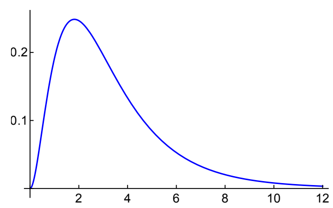
Mathematica: BetaDistribution[a, b]

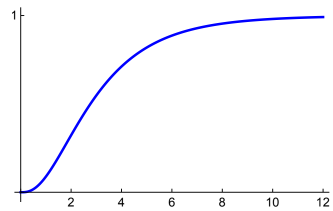
$I_{(a,b)}(x) =$ HP: fisher_cdf($2a, 2b, \frac{bx}{a(1-x)}$)
CASIO: F » FCd→FCD($0, \frac{bx}{a(1-x)}, 2a, 2b$)
TI: Fcdf→ Fcdf($0, \frac{bx}{a(1-x)}, 2a, 2b$)

$I^{-1}_{(a,b)}(p) =$ HP: $1/(1 + \frac{b}{a\cdot \text{fisher\_icdf}(2a,2b,p)})$
CASIO: F » InvF→ $1/(1 + \frac{b}{a\cdot \text{InvFCD}(1-p,2a,2b)})$
TI (CX): InvF→ $1/(1 + \frac{b}{a\cdot \text{InvF}(p,2a,2b)})$
TI (83/84): MATH » Solver: Fcdf($0, \frac{bx}{a(1-x)}, 2a, 2b$) $- p = 0$ » ⇓ » $X = 0.5$ » Alpha » Enter

Approximations: normal approximation (7.2) when $a, b > 10$.

7.7.3 GAMMA-GAMMA DISTRIBUTION, $g\gamma_{(k,\kappa,\tau)}(x)$, $x \in (0, \infty)$





pdf: $T \sim g\gamma_{(k,\kappa,\tau)}(t) = \frac{1}{\mathrm{B}_{(k,\kappa)}} \cdot \frac{\tau^{\kappa} \cdot t^{k-1}}{(\tau+t)^{\kappa+k}}$

CDF: $G\Gamma_{(k,\kappa,\tau)}(t) = F_{(2k,2\kappa)}\left(\frac{\kappa t}{k\tau}\right)$

$\mu_T = \frac{k\tau}{\kappa-1}$

$\sigma_T^2 = \frac{k\tau^2(\kappa+k-1)}{(\kappa-2)(\kappa-1)^2}$

$\widetilde{T} = G\Gamma^{-1}_{(k,\kappa,\tau)}(0.5)$

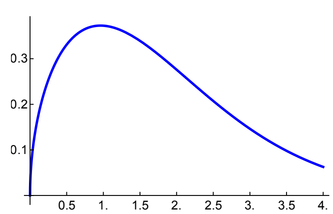
$T_{max} = \frac{k-1}{\kappa+1}\tau$

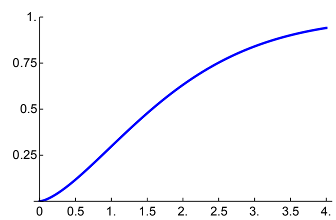
Mathematica: BetaPrimeDistribution[k, κ, τ]

$G\Gamma_{(k,\kappa,\tau)}(t) =$ HP: fisher_cdf($2k, 2\kappa, \frac{\kappa t}{k\tau}$)
CASIO: F » FCd→FCD($0, \frac{\kappa t}{k\tau}, 2k, 2\kappa$)
TI: Fcdf→ Fcdf($0, \frac{\kappa t}{k\tau}, 2k, 2\kappa$)

$G\Gamma^{-1}_{(k,\kappa,\tau)}(p) =$ HP: $\frac{k\tau}{\kappa}$ ∗ fisher_icdf($2k, 2\kappa, p$)
CASIO: F » InvF→ $\frac{k\tau}{\kappa}$ ∗ InvFCD($1-p, 2k, 2\kappa$)
TI (CX): InvF→ $\frac{k\tau}{\kappa}$ ∗ InvFCD($p, 2k, 2\kappa$)
TI (83/84): MATH » Solver: Fcdf($0, \frac{\kappa x}{k\tau}, 2k, 2\kappa$) $- p = 0$ » ⇓ » $X = 0.5$ » Alpha » Enter

7.7.4 WEIBULL DISTRIBUTION, $weib_{(k,\lambda)}(x)$, $x \in (0, \infty)$





pdf: $X \sim \mathrm{weib}_{(\lambda,k)}(x) = \frac{k}{\lambda} \cdot \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}$

CDF: $\mathrm{WEIB}_{(\lambda,k)}(x) = 1 - e^{-\left(\frac{x}{\lambda}\right)^k}$

$\mu_X = \frac{\lambda}{k} \cdot \Gamma\left(\frac{1}{k}\right)$

$\sigma_X^2 = \lambda^2 \left(\Gamma\left(1+\frac{2}{k}\right) - \left(\Gamma\left(1+\frac{1}{k}\right)\right)^2\right)$

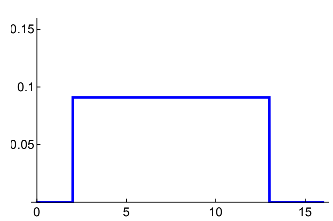
$\widetilde{X} = k\sqrt[\lambda]{\ln 2}$

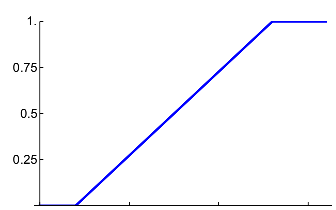
$X_{max} = k\sqrt[\lambda]{1-\frac{1}{\lambda}}$

Special case: The Rayleigh distribution is $rayl_{(\sigma)}(x) = weib_{(2,\sqrt{2}\cdot\sigma)}(x)$.

Mathematica: WeibullDistribution[k , λ]
Calculator: Use the formulas above.

7.7.5 UNIFORM DISTRIBUTION, $uni_{(a,b)}(x)$, $x \in (a, b)$





pdf: $X \sim \mathrm{uni}_{(a,b)}(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$

CDF: $\mathrm{UNI}_{(a,b)}(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a,b] \\ 1 & x \geq b \end{cases}$

$\mu_X = E[X] = \frac{a+b}{2}$

$\sigma_X^2 = \frac{(b-a)^2}{12}$

$\widetilde{X} = \frac{a+b}{2}$

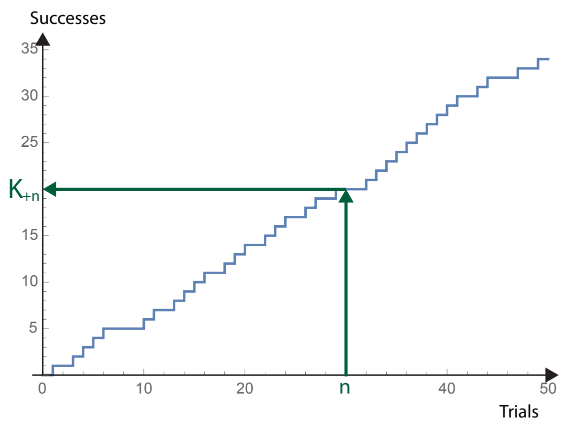
$X_{max} = [a,b]$ (the interval)

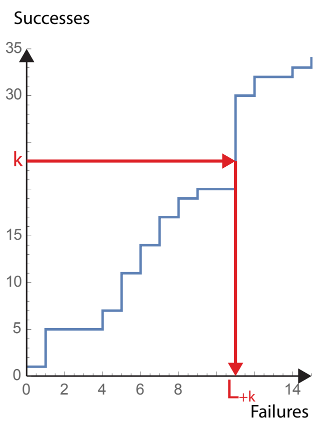
Mathematica: UniformDistribution[a, b]
Calculator: Use the formulas above.

8 Processes

8.1. Bernoulli process with parameter p

A Bernoulli process with parameter p is a sequence of independent Bernoulli distributed stochastic variables $X_1, X_2, \ldots$, where each $X_j \sim bern_p = bin_{1,p}$.





8.1.1 Relevant distributions: Binomial w/special case Bernoulli (6.1.2), Negative binomial (6.1.4), Beta (7.7.2), F (7.7.1), Beta binomial (6.1.5), Beta negative binomial (6.1.6)

8.1.2 Elementary calculations

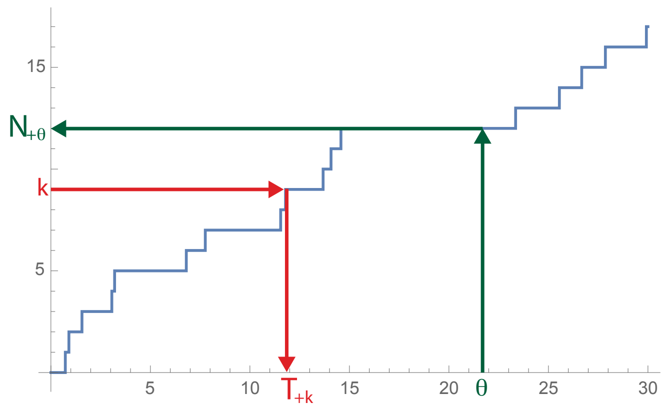
	$E[p]$	Successes in n trials	$E[K_{+n}]$	Errors before k successes	$E[L_{+k}]$
Known p	p	$K_{+n} \sim \text{bin}_{(n,p)}(x)$ (6.1.2)	np	$L_{+k} \sim \text{nb}_{(k,p)}(x)$ (6.1.4)	$\frac{k(1-p)}{p}$
Unknown p, $p \sim \beta_{(a,b)}(t)$	$\frac{a}{a+b}$	$K_{+n} \sim \beta\text{b}_{(a,b,n)}(x)$ (6.1.5)	$\frac{na}{a+b}$	$L_{+k} \sim \beta\text{nb}_{(a,b,k)}(x)$ (6.1.6)	$\frac{kb}{a-1}$

8.1.3 Addition rules

- For known p, the following applies to independent X, Y:
  - If $X \sim \text{bin}_{(n,p)}(x)$ and $Y \sim \text{bin}_{(m,p)}(x)$, then $X + Y = Z \sim \text{bin}_{(m+n,p)}(x)$
  - If $X \sim \text{nb}_{(k,p)}(x)$ and $L_{+l} \sim \text{nb}_{(l,p)}(x)$, then $X + Y = Z \sim \text{nb}_{(k+l,p)}(x)$
- For unknown $p \sim \beta_{(a,b)}$ we have the weaker addition rules:
  - $K_{+n} \sim \beta\text{b}_{(a,b,n)}(x)$ and $K_{+m} \sim \beta\text{b}_{(a,b,m)}(x)$, and also $K_{+(m+n)} \sim \beta\text{b}_{(a,b,m+n)}(x)$
  - $L_{+k} \sim \beta\text{nb}_{(a,b,k)}(x)$ and $L_{+l} \sim \beta\text{nb}_{(a,b,l)}(x)$, and also $L_{+(k+l)} \sim \beta\text{nb}_{(a,b,k+l)}(x)$

8.2. Poisson process with parameter λ

A Poisson process with parameter λ is a sequential observation of independent occurences of $\top$. The process is ruled by the rate parameter λ, indicating the expected number of successes per (time) unit.



8.2.1 Relevant distributions: Gamma (7.6), Poisson (6.1.3), F (7.7.1), Gamma-gamma (7.7.3), Negative binomial (6.1.4)

8.2.2 Elementary calculations

	$E[\lambda]$	Successes during θ units	$E[N_{+\theta}]$	Waiting units until k successes	$E[T_{+k}]$
Known λ	λ	$N_{+\theta}$ ~$\text{pois}_{\lambda\theta}(x)$ (6.1.3)	$\lambda\theta$	$T_{+k} \sim \gamma_{(k,\lambda)}(x)$ (7.6)	$\frac{k}{\lambda}$
Unknown λ, $\lambda \sim \gamma_{(\kappa,\tau)}(t)$	$\frac{\kappa}{\tau}$	$N_{+\theta}$ ~$\text{nb}_{\left(\kappa,\frac{\tau}{\tau+\theta}\right)}(x)$ (6.1.5)	$\frac{\kappa\theta}{\tau}$	T_{+k} ~$\text{g}\gamma_{(k,\kappa,\tau)}(x)$ (7.7.3)	$\frac{k\tau}{\kappa-1}$

8.2.3 Addition rules

- For known λ, we have the following addition rules when X, Y are independent:
  - If $X \sim \gamma_{(k,\lambda)}(x)$ and $Y \sim \gamma_{(l,\lambda)}(x)$, then $X + Y = Z \sim \gamma_{(k+l,\lambda)}(x)$
  - If X ~$\text{pois}_{\lambda\cdot\theta_1}(x)$ and Y ~$\text{pois}_{\lambda\cdot\theta_2}(x)$, then $X + Y = Z$ ~$\text{pois}_{\lambda\cdot(\theta_1+\theta_2)}(x)$
  - General: If X ~$\text{pois}_{\lambda_1}(x)$ and Y ~$\text{pois}_{\lambda_2}(x)$, then $X + Y = Z$ ~$\text{pois}_{\lambda_1+\lambda_2}(x)$
- For unknown $\lambda \sim \gamma_{(\kappa,\tau)}$ we have the weaker addition rules:
  - T_{+k} ~$\text{g}\gamma_{(k,\kappa,\tau)}(x)$ and T_{+l} ~$\text{g}\gamma_{(l,\kappa,\tau)}(x)$, and also $T_{+(k+l)}$ ~$\text{g}\gamma_{(k+l,\kappa,\tau)}(x)$
  - $N_{+\theta_1}$ ~$\text{nb}_{\left(\kappa,\frac{\tau}{\tau+\theta_1}\right)}(x)$ and $N_{+\theta_2}$ ~$\text{nb}_{\left(\kappa,\frac{\tau}{\tau+\theta_2}\right)}(x)$, and also $N_{+(\theta_1+\theta_2)}$ ~$\text{nb}_{\left(\kappa,\frac{\tau}{\tau+\theta_1+\theta_2}\right)}(x)$

8.3. Gaussian process with parameters μ and σ

A Gaussian process with parameters μ and σ is a sequence of independent stochastic variables $X_1, X_2, \ldots$ where each $X_j \sim \phi_{(\mu,\sigma)}$.

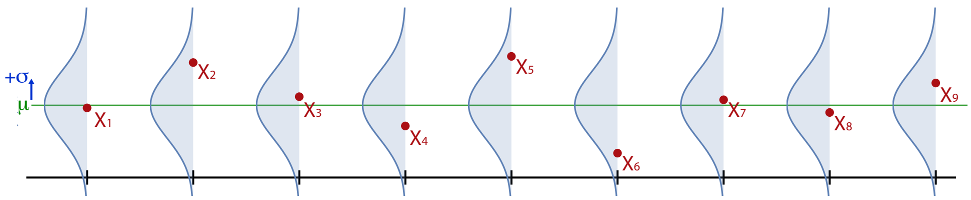


Figure 8.1: $X_1, X_2, \ldots$ where each $X_j \sim \phi_{(\mu,\sigma)}$; μ and σ as marked on the diagram.

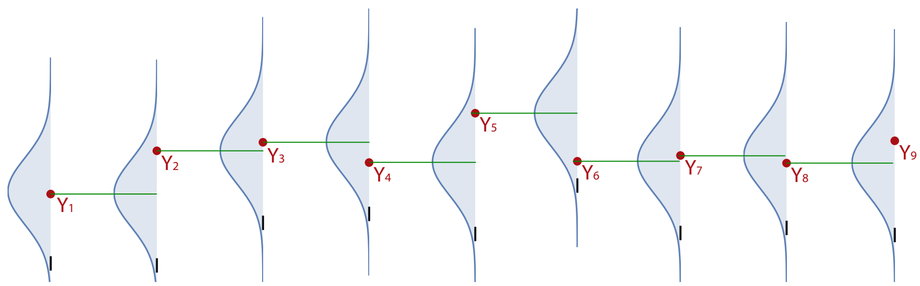


Figure 8.2: A cumulative process $Y_1, Y_2, \ldots$ where $Y_k = X_1 + X_2 + \cdots + X_k$, and $\mu = 0$

Relevant distributions: *normal* ϕ (7.1), t *distribution* (7.4), *gamma* γ (7.6), F distribution (7.7.1)

8.3.1 Rules

- When $X_j \sim \phi_{(\mu,\sigma)}$, and $Y_k = X_1 + X_2 + \cdots + X_k$, then $Y_k \sim \phi_{(k\cdot\mu\,,\,\sqrt{k}\cdot\sigma)}$
- When $X_j \sim \phi_{(\mu,\sigma)}$, and you have $Y_k = \max\{X_1, X_2, \ldots, X_k\}$ and $Z_k = \min\{X_1, X_2, \ldots, X_k\}$, let $p_x = \Phi_{(mu\,,\,\sigma)}(x)$. Then $P(Y_k \leq x) = p_x^k$, and $P(Z_k \leq x) = 1 - (1 - p_x)^k$.
- Given $\{X_1, X_2, \ldots, X_n\}$, where $X_j \sim \phi_{(\mu,\sigma)}$, let $p_x = \Phi_{(\mu\,,\,\sigma)}(x)$. The probability that precisely k of the X are less than x, is given by $\text{bin}_{(n,p_x)}(k)$, and the probability that k or fewer of the $X$are less than x is given by $\text{BIN}_{(n,p_x)}(k)$.

9 Multivariate statistics

DATA: Formulas for data pairs $\{(x_1, y_1), \ldots, (x_n, y_n)\}$

See (2.2).

9.1. PROBABILITY: Multivariate probability distributions

9.1.1 Cumulative distribution: $F_{XY}(a,b) = P(X \le a,\, Y \le b) = \begin{cases} \displaystyle\sum_{x \le a,\ y \le b} f_{XY}(x,y) \\ \displaystyle\int_{-\infty}^{a} \int_{-\infty}^{b} f_{XY}(x,y)\,dy\,dx \end{cases}$

9.1.2 Cumulative marginal distribution: $F_X(a) = F_{XY}(a, \infty) = \begin{cases} \displaystyle\sum_{x \le a,\ \text{all } y} f_{XY}(x,y) \\ \displaystyle\int_{-\infty}^{a} \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy\,dx \end{cases}$

9.1.3 Distribution: $f_{XY}(a,b) = \begin{cases} F_{XY}(a,b) - F_{XY}(a-1,b) - F_{XY}(a,b-1) + F_{XY}(a-1,b-1) \\ \frac{\partial^2}{\partial y \partial x} F_{XY}(x,y) \end{cases}$

9.1.4 Marginal probability distribution: $f_X(a) = \begin{cases} \displaystyle\sum_{\text{all } y} f_{XY}(a,y) = F_X(a) - F_X(a-1) \\ \frac{\partial}{\partial y} F_X(a) = \int_{-\infty}^{\infty} f_{XY}(a,y)\,dy \end{cases}$

9.1.5 Probability: $P(a < X < c,\, b < Y < d) = F_{XY}(c,d) - F_{XY}(a,d) - F_{XY}(c,b) + F_{XY}(a,b)$

9.1.6 $E[XY] = \begin{cases} \displaystyle\sum_{\text{all } x,y} xy \cdot p_{xy} & \text{(discrete)} \\ \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f(x,y) dy dx & \text{(continuous)} \end{cases}$

9.1.7 Covariance: $\sigma_{XY} = E[XY] - \mu_X \mu_Y$

9.1.8 Correlation: $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$

10 Inference (General Bayesian)

10.1. Bayes' theorem with discrete prior

10.1.1 Probability that the (next) observation is B: Given (*posterior*) probability function P_n, we get $P_n(B)$, the probability that the (next) observation is B, from the following table:

k	(1) $P_n(A_k)$	(2) $P_n(B\|A_k)$	(3) $P_n(A_kB) = P_n(A_k) \cdot P_n(B\|A_k)$
			(4 - answer) $P_n(B) = \sum_j P_n(A_jB)$

10.1.2 Bayes' theorem, basic version: If we partition Ω into disjoint (mutually exclusive) alternatives $A_1, A_2, \ldots$, and observe B, the the probabilities of the A_k are updated as follows:

$$P(A_k|B) = \frac{P(A_k) \times P(B|A_k)}{\sum_j P(A_j) \times P(B|A_j)}$$

10.1.3 Bayes' theorem, tabular notation:

Alt.	Prior	Likelihood	Joint probabillity	Posterior
A_k	(1) $P_n(A_k)$	(2) $P_n(B\|A_k)$	(3) $P_n(A_kB) = P_n(A_k) \cdot P_n(B\|A_k)$	(5 - answer) $P_{n+1}(A_k\|B) = P_n(A_k\|B) = \frac{P_n(A_kB)}{P_n(B)}$
		Total probability:	(4) $P_n(B) = \sum_{j=1}^{n} P_n(A_jB)$	

10.2. Bayes' theorem, function version

10.2.1 Bayes' theorem for probability distributions f_n:

	Prior	Likelihood	Joint probabillity	Posterior
x	(1) $f(x) = f_n(x)$	(2) $g(k) = h_x(y)$	(3) $f(x) \cdot g(x)$	(5 - answer) $f_{n+1}(x) = \frac{f(x) \cdot g(x)}{S}$
		Total probability:	(4) $S = \sum_x f(x) \cdot g(x)$	(discrete prior)
			(4) $S = \int_{-\infty}^{\infty} f(x) \cdot g(x) dx$	(continuous prior)

where $h_x(y)$ is the conditional probability (density) for the observation y, given thatat $X = x$.

10.3. Repeat update

10.3.1 Upon a new observation B_{n+1} (set theory version) or y_{n+1} (function version):
Repeat the procedure, and use the previous *posterior* f_n / P_n as the new *prior* to find a new *posterior* f_{n+1} / P_{n+1}.

11 General estimates

The starting point of the section is a probability distribution $g(x)$, and a stochastic variable Θ (typically a parameter or a predictive variable) with $\Theta \sim g(x)$, cumulative distribution $G(x)$ and inverse cumulative distribution $G^{-1}(p)$.

11.1. Point estimates for $\Theta \sim g(x)$

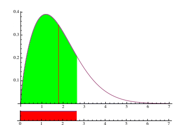
11.1.1 Median: $\tilde{\Theta} = G^{-1}(0.5)$

11.1.2 Expected value: $E[\theta] = \mu_\Theta = \int_D t \cdot g(t)dt$ where D are the possible values for Θ.

11.1.3 Mode: $\Theta_{MAP} = \Theta_{max}$ is the t value that maximizes $g(t)$.

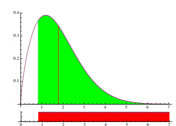
11.2. Interval estimates

We typically write I^Θ for the interval estimate ("credible interval") when Θ is a parameter, but I^+ for the interval estimate ("predictive interval") when Θ is a next observation. Below, we write them both as I^Θ.



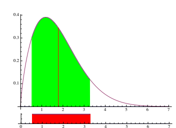
11.2.1 1-sided left $(1-\alpha)100\%$ interval estimate $I_{\alpha,l}$, given a probability distribution $g(x)$:

$$I^\Theta_{\alpha,l} = \left(G^{-1}(0), G^{-1}(1-\alpha)\right)$$



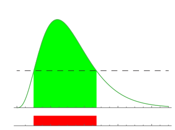
11.2.2 1-sided right $(1-\alpha)100\%$ interval estimate $I_{\alpha,r}$, given a probability distribution $g(x)$:

$$I^\Theta_{\alpha,r} = \left(G^{-1}(\alpha), G^{-1}(1)\right)$$



11.2.3 2-sided $(1-2\alpha)100\%$ interval estimate I_α, given a probability distribution $g(x)$:

$$I^\Theta_{\alpha} = \left(G^{-1}(\alpha), G^{-1}(1-\alpha)\right)$$



11.2.4 HPD interval H^Θ_l of width l, given a probability distribution $g(x)$:

H^Θ_l is the interval $(a, a+l)$

where a is the value maximizing $H(a) = G(a+l) - G(a)$. At this value, $g(a+l) = g(a)$.

12 Comparison and hypothesis testing

12.1. Comparison

12.1.1 Utility function: $u_A(\theta)$ is the utility of an alternative H_A, given a value of θ.

12.1.2 The expected utility of an alternative H_A, given that $\Theta \sim g(\theta)$, is $U_\theta = E[u_A(\Theta)] = \int_{-\infty}^{\infty} u_A(\theta) \cdot g(\theta) d\theta$

Special cases:

1. Linear utility: $u_A(\theta) = a + b\theta$ gives $U_A = a + b \cdot E[\Theta]$
2. Two-value utility: $u_B(\theta) = \begin{cases} a & \theta < \theta_0 \\ b & \theta > \theta_0 \end{cases}$ given that $U_B = a \cdot P(\Theta < \theta_0) + b \cdot P(\Theta > \theta_0)$

12.1.3 Utility maximizing choice: Given $\Theta \sim g(\theta)$, and two alternatives H_A and H_B with utility functions $u_A(\theta)$ and $u_B(\theta)$, the utility maximizing choice is the alternative with the largest expected utility.

12.1.4 Two-value utility maximizing choice: Given $\Theta \sim g(\theta)$, a critical value θ_0, and two alternatives H_A and H_B where $u(\theta) = u_A(\theta) - u_B(\theta)$ is a two-value utility function,

$$u(\theta) = \begin{cases} w_A & \theta < \theta_0 \\ -w_B & \theta > \theta_0 \end{cases}$$

with $w_A, w_B > 0$. Then the alternatives correspond to H_A: $\Theta < \theta_0$, and H_B: $\Theta > \theta_0$. The utility maximizing choice is then

$$\left.\begin{array}{lll} A & \text{if} & w_A \cdot P(A) > w_B \cdot P(B) \\ B & \text{if} & w_A \cdot P(A) < w_B \cdot P(B) \end{array}\right\} \begin{cases} P(A) = P(\Theta < \theta_0) \\ P(B) = P(\Theta > \theta_0) \end{cases}$$

12.2. Bayesian hypothesis test

12.2.1 Hypothesis test with significance α tests the null hypothesis / conservative hypothesis H_0 against the alternative / daring hypothesis H_1. The significance is set in advance, independently of the data (see 12.2.2 below for a suggestion how to determine α).

- If H_0: $\Theta \leq \theta_0$, then H_1: $\Theta > \theta_0$.
- If H_0: $\Theta \geq \theta_0$, then H_1: $\Theta < \theta_0$.

Decisions are then made in this fashion:

- If $P(H_0) \geq \alpha$, the utility maximizing choice is H_0. (formally, we say we "*don't reject* H_0")
- If $P(H_0) < \alpha$, the utility maximizing choice is H_1. (formally, we say we "*reject* H_0")

Hypothesis test is primarily a frequentist way of formulating utility maximizing choices, so we translate utility maximizing choices into hypothesis testing in the Bayesian context as follows:

12.2.2 Reformulating utility maximizing choice as hypothesis test: Given the alternatives H_A and H_B in 12.1.4, call the alternative with maximal w value the *Null* hypothesis H_0, and let the *test significance* be $\alpha = \frac{w_1}{w_0+w_1}$, where w_0 is the largest of w_A and w_B, and w_1 is the smallest. The *alternative hypothesis* H_1 is the alternative with the smallest w value. The case $\Theta = \theta_0$ is included in H_0. Significance α and hypotheses H_0 and H_1 are often given directly without explicit reference to maximizing choice utility.

13 Bayesian inference for Bernoulli processes

13.1. Bayes' theorem for Bernoulli processes ("The Bernoulli version")

13.1.1 Inference for hyperparameters for the Bernoulli parameter p

Prior hyperparameters:	Observations:	*Posterior* hyperparameters
$P_0 \models a_0$	$k =$ number of $\top$	$P_1 \models a_1 = a_0 + k$
b_0	$l =$ number of $\bot$	$b_1 = b_0 + l$

Posterior values:

1. $p \sim \beta_{(a_1\,,\,b_1)}$ - *posterior* probability distribution for p.
2. $K_{+m} \sim \beta b_{(a_1\,,\,b_1\,,\,m)}$ - number of $\top$ during the next m observations.
3. $L_{+s} \sim \beta nb_{(a_1\,,\,b_1\,,\,s)}$ - number of new $\bot$ before s new $\top$.

which also gives us $p_1 = E[p] = \frac{a_1}{a_1+b_1}$.

A special case is if we started out with Novick & Hall's prior, and our posterior hyperparameters are that either a_1 or b_1 are equal to 0. Then p has a discrete probability distribution whether either $P(\top) = 0$ or $(\top) = 1$, with the opposite probability assigned to $\bot$.

13.2. Prior hyperparameters for Bernoulli processes

13.2.1 Neutral prior hyperparameters are when $a_0 = b_0 = u \in [0, 1]$, and either

1. $u = 0$: "Total ignorance" (*Novick & Hall, Haldane, Jaynes*). (If you don't even know whether both $\top$ and $\bot$ are *possible.*)
2. $u = 0.5$: "Ordinary ignorance" (*Jeffreys*). (You know that both $\top$ and $\bot$ are possible, but nothing more.)
3. $u = 1$: "Informed ignorance" (*Laplace, Bayes*). (If your starting point is a symmetry between $\top$ and $\bot$.)

If, in a concrete case, you are unable to choose between two or three of these possibilities: choose the smallest u under consideration!

13.2.2 Informative prior hyperparameters are one of these:

1. Prior hyperparameters = *Posterior* hyperparameters from the previous update.
2. Let p_0 be your estimate of p, and let κ_0 be the number of observations required to have your degree of certainty. Your hyperparameters are then $a_0 = \kappa_0 p_0$ and $b_0 = \kappa_0(1 - p_0)$. Note: Informative priors require $a_0, b_0 \geq 1$.

13.3. Comparison and hypothesis test

For comparison of p against a fixed reference value p_0, see the general formulas for comparison and hypothesis testing in chapter 12.

13.3.1 Comparison: Given two independent stochastic variables with *posterior* distributions $\psi \sim \beta_{(a,b)}$ and $\pi \sim \beta_{(\theta,\rho)}$, then (see 1.3.10 for the function B) we have

Exact:

$$P(\psi \leq \pi) = \sum_{k=0}^{\theta-1} \frac{B(a+k\,,\,b+\rho)}{(\rho+k)\cdot B(k+1\,,\,\rho)\cdot B(a,b)} = \sum_{k=0}^{\theta-1} \frac{\binom{a+b}{a}\cdot\binom{k+\rho}{k}\cdot ab\rho(a+b+k+\rho)}{\binom{a+b+k+\rho}{a+k}\cdot(a+b)(k+\rho)(a+k)(b+\rho)}$$

Approximation (using normal approximation); good when $a, b, \theta, \rho > 10$)

$$P(\psi \leq \pi) \approx \Phi_{\left(\frac{a}{a+b}-\frac{\theta}{\theta+\rho}\,,\,\sqrt{\frac{ab}{(a+b)^2(a+b+1)}+\frac{\theta\rho}{(\theta+\rho)^2(\theta+\rho+1)}}\right)}(0)$$

13.4. Estimates

See chapter 11 for the general formulas for point and interval estimates (11.1, 11.2), and use formulas from the relevant chapters for the *posterior* or *predictive* distributions.

13.4.1 HPD interval of width l for $p \sim \beta_{(a,b)}$

$$H_l^p = (k, k+l)$$

where k is the real-valued solution of $(k+l)^{a-1}(1-k-l)^{b-1} - k^{a-1}(1-k)^{b-1} = 0$

13.4.2 Sample size n for HPD interval H_l^p for $p \sim \beta_{(a,b)}$ with $a, b > 1$ such that $P(p \in H_l^p) \geq 1 - 2\alpha$, and the width of $I_{2\alpha}^p$ is less than or equal to l, is

$$n = \frac{z_\alpha^2}{l^2} - a - b$$

14 Bayesian inference for Poisson processes

14.1. Bayes' theorem for Poisson-processes ("The Poisson version")

14.1.1 Inference for hyperparameters for the Poisson rate parameter λ.

Prior hyperparameters:	Observations:	*Posterior* hyperparameters
$P_0 \models \kappa_0$	$n =$ number of occurences	$P_1 \models \kappa_1 = \kappa_0 + n$
τ_0	$t =$ number of (time) units (time, or number of tries)	$\tau_1 = \tau_0 + t$

Posterior values:

1. $\lambda \sim \gamma_{(\kappa_1\,,\,\tau_1)}(l)$ - *posterior* probability distribution for λ.
2. $N_{+\theta} \sim nb_{(\kappa_1\,,\,\frac{\tau_1}{\tau_1+\theta})}(\eta)$ - occurences in the next θ (time) units.
3. $T_{+k} \sim g\gamma_{(k\,,\,\kappa_1\,,\,\tau_1)}(t)$ - waiting time for the next k occurences.

14.2. Prior hyperparameters

14.2.1 Neutral prior hyperparameters are $\kappa_0 = \tau_0 = 0$.

14.2.2 Informative prior hyperparameters are either of the following:

1. Prior hyperparameters = *Posterior* hyperparameters from a previous update.
2. Let λ_0 be your estimate of λ, and let κ_0 be the number of occurences required to have your degree of certainty. Alternatively, instead of κ_0, let τ_0 be the number of (time) units required to have your degree of certainty. The two values are connected via the formula $\lambda_0 = \frac{\kappa_0}{\tau_0}$.

14.3. Comparison and hypothesis test

For comparison of λ against a fixed reference value λ_0, see the general formulas for comparison and hypothesis testing in chapter 12.

14.3.1 Comparison: Given two independent stochastic variables with *posterior* distributions

$$\begin{aligned} A &\sim \gamma_{(k,l)}(t) \\ B &\sim \gamma_{(m,n)}(t) \end{aligned}$$

then

$$P(A < B) = F_{(2k,2m)}\left(\frac{ml}{kn}\right)$$

14.4. Estimates

for λ with *posterior* hyperparameters κ and τ, and probability distribution $\lambda \sim \gamma_{(\kappa\,,\,\tau)}$

See chapter 11 for the general formulas for point and interval estimates (11.1, 11.2), and use formulas from the relevant chapters for the *posterior* or *predictive* distributions.

14.4.1 HPD interval of width b for λ:

$$H_b^\lambda = (a, a+b)$$

where

$$a = \frac{b}{e^{\frac{\tau b}{\kappa - 1}} - 1}$$

14.4.2 Sample size n for HPD interval H_b^λ for λ with *prior* hyperparameter κ_0, such that the relative interval width, $r = \frac{b}{E[\lambda]}$, is less than a fixed value R:

$$n \geq \frac{4z_\alpha^2}{R^2} - \kappa_0$$

15 Bayesian inference for gaussian processes

15.1. Bayes' theorem for Gaussian processes ("Gaussian version")

for estimating the Gaussian parameters μ and $\tau = \frac{1}{\sigma^2}$.

15.1.1 Inference for hyperparameters given n measurements $x_1, \ldots, x_n$ with (see section 2.2)

- sum $\Sigma_x = \sum_{k=1}^{n} x_k$ and mean $\bar{x} = \frac{\Sigma_x}{n}$
- sum of squared deviances from average $SS_x = \sum_{k=1}^{n} (x_k - \bar{x})^2$
- sum of squared deviances from mean $SB_x = \sum_{k=1}^{n} (x_k - m_0)^2 = SS_x + n \cdot (\bar{x} - m_0)^2$

Prior hyperparameters: next page.

	$\sigma = s_0$ (known)	σ unknown
$\mu = m_0$ (known)	No update of hyperparameters. Posterior values: $\tau = 1/s_0^2$ $\mu = m_0$ $X_+ \sim \phi_{(m_0,s_0)}(x)$	*Posterior* hyperparameters: $\left.\begin{array}{rcl}\nu_1 &=& \nu_0 + n \\ SS_1 &=& SS_0 + SB_x\end{array}\right\} \quad s_1^2 = \frac{SS_1}{\nu_1}$ Posterior values: $\tau \sim \gamma_{(\frac{\nu_1}{2}, \frac{B_1}{2})}(t)$ $\mu = m_0$ $X_+ \sim t_{(m_0, s_1, \nu_1)}(x)$
μ unknown	*Posterior* hyperparameters: $\left.\begin{array}{rcl}\kappa_1 &=& \kappa_0 + n \\ \Sigma_1 &=& \Sigma_0 + \Sigma_x\end{array}\right\} \quad m_1 = \frac{\Sigma_1}{\kappa_1}$ Posterior values: $\tau = 1/s_0^2$ $\mu \sim \phi_{\left(m_1, s_0\sqrt{\frac{1}{\kappa_1}}\right)}(x)$ $X_+ \sim \phi_{\left(m_1, s_0\sqrt{1+\frac{1}{\kappa_1}}\right)}(x)$	*Posterior* hyperparameters: $\left.\begin{array}{rcl}\kappa_1 &=& \kappa_0 + n \\ \Sigma_1 &=& \Sigma_0 + \Sigma_x\end{array}\right\} \quad m_1 = \frac{\Sigma_1}{\kappa_1}$ $\left.\begin{array}{rcl}\nu_1 &=& \nu_0 + n \\ C_1 &=& C_0 + \Sigma_{x^2} \\ SS_1 &=& C_1 - \kappa_1 \cdot m_1^2\end{array}\right\} \quad s_1^2 = \frac{SS_1}{\nu_1}$ $SS_1 = SS_0 + SS_x + n \cdot \frac{\kappa_0}{\kappa_1}(x - m_0)^2$ (shortcut) Posterior values: $\tau \sim \gamma_{(\frac{\nu_1}{2}, \frac{B_1}{2})}(t)$ $\mu \sim t_{\left(m_1, s_1 \cdot \sqrt{\frac{1}{\kappa_1}}, \nu_1\right)}(x)$ $X_+ \sim t_{\left(m_1, s_1 \cdot \sqrt{1+\frac{1}{\kappa_1}}, \nu_1\right)}(x)$

15.2. *Prior* hyperparameters for Gaussian processes

15.2.1 We have $3\frac{1}{2}$ possibilities for μ:

1. μ is fully known: $\mu = m_0$
2. μ partially known (informative prior): Let m_0 be your estimate of μ, and let κ_0 be the number of observations required to have your degree of certainty. Let $\Sigma_0 = m_0\kappa_0$.

$2\frac{1}{2}$. μ partially known (informative prior), with prior $\mu \sim \phi_{(m_{pre}, s_{pre})}$, and $\sigma = s_0$ is known. Then, $\kappa_0 = \frac{s_0^2}{s_{pre}^2}$, and $\Sigma_0 = m_{pre}\kappa_0$.

3. μ is fully unknown (neutral prior): $\kappa_0 = 0$ and $\Sigma_0 = 0$.

15.2.2 We have 3 possibilities for $\tau = \frac{1}{\sigma^2}$:

1. $\tau = \frac{1}{\sigma^2}$ is fully known: $\tau = \tau_0 = \frac{1}{s_0^2}$
2. τ partially known (informative prior): Let s_0 be your estimate of σ, and let n_0 be the number of observations required to have your degree of certainty. Let $\nu_0 = n_0 - 1$ and $SS_0 = s_0^2 \cdot \max(0, \nu_0)$, and let $C_0 = SS_0 + \kappa_0 \cdot m_0$
3. $\tau = \frac{1}{\sigma^2}$ is fully unknown (neutral prior): Let $\nu_0 = -1$ and $SS_0 = 0$, and let $C_0 = SS_0 + \kappa_0 \cdot m_0^2$

15.3. Comparison of a parameter against a fixed value

For comparison of μ and $\tau = \frac{1}{\sigma^2}$ against a fixed reference value m_0 and $\tau_0 = \frac{1}{s_0^2}$, see the general formulas for hypothesis testing and comparisons in chapter 12.

15.3.1 Comparing $\tau = \frac{1}{\sigma^2}$, with *posterior* distribution $\tau \sim \gamma_{(k,l)}(t)$ against the fixed value $\tau_0 = \frac{1}{\sigma_0^2}$:

$$P(\sigma \geq \sigma_0) = P(\tau \leq \tau_0) = \Gamma_{(k,l)}(\tau_0)$$

15.3.2 (Known σ) Comparing $\mu$with *posterior* distribution $\mu \sim \phi_{(m,s)}(x)$ against the fixed value μ_0:

$$P(\mu \leq \mu_0) = \Phi_{(m,s)}(\mu_0)$$

15.3.3 (Unknown σ) Comparing $\mu$with *posterior* distribution $\mu \sim t_{(m,s,n)}(x)$ against the fixed value μ_0:

$$P(\mu \leq \mu_0) = T_{(m,s,n)}(\mu_0)$$

15.4. Comparing two parameters

15.4.1 Comparison Given two independent stochastic variables with *posterior* distributions $\tau_1 \sim \gamma_{(k,l)}(t)$ and $\tau_2 \sim \gamma_{(m,n)}(t)$,

$$P(\sigma_1 > \sigma_2) = P(\tau_1 < \tau_2) = F_{(2k,2m)}\left(\frac{ml}{kn}\right)$$

15.4.2 Comparison of two independent μ_n when σ is known, with *posterior* distributions $\mu_1 \sim \phi_{(m_1,s_1)}(x)$ and $\mu_2 \sim \phi_{(m_2,s_2)}(x)$: Use formula (7.3) to find the distribution $\phi_{(m,s)}(x)$ for $Z = \mu_1 - \mu_2$. Then,

$$P(\mu_1 < \mu_2) = \Phi_{(m,s)}(0)$$

15.4.3 Comparison of two independent μ_n when σ is unknown, with *posterior* distributions $\mu_1 \sim t_{(m_1,s_1,\nu_1)}(x)$ and $\mu_2 \sim t_{(m_2,s_2,\nu_2)}(x)$: Use formula (7.5) to find the distribution $t_{(m,s,\nu)}(x)$ for $Z = \mu_1 - \mu_2$. Then,

$$P(\mu_1 < \mu_2) = T_{(m,s,\nu)}(0)$$

15.5. Estimates for μ

See chapter 11 for the general formulas for point and interval estimates (11.1, 11.2), and use formulas from the relevant sections for the *posterior* or *predictive* distributions.

15.5.1 Sample size n, such that the width of the $(1-2\alpha)100\%$ symmetrical credible interval is less than l, given known $\sigma = s_0$ and *prior* hyperparameter κ_0:

$$n = \frac{4z_\alpha^2}{l^2} \cdot s_0^2 - \kappa_0$$

15.5.2 Sample size n, such that the width of the $(1-2\alpha)100\%$ symmetrical credible interval is less than l, with unknown σ, and *prior* hyperparameters κ_0, ν_0 and B_0:

$$n = \frac{4t_{2\nu_0,\alpha}^2}{l^2} \cdot \frac{B_0}{\nu_0} - \kappa_0$$

15.6. Point estimates for $\tau = \frac{1}{\sigma^2}$,

See chapter 11 for the general formulas for point and interval estimates (11.1, 11.2), and use formulas from the relevant sections for the *posterior* or *predictive* distributions.

15.6.1 HPD interval of width l for $\tau \sim \gamma_{(k,\lambda)}$

$$H_l^\tau = (a, a+l)$$

where

$$a = \frac{l}{e^{\frac{\lambda l}{k-1}} - 1}$$

15.6.2 Sample size n for HPD interval H_b^τ with *prior* hyperparameter ν_0, such that the relative interval width, $r = \frac{b}{E[\lambda]}$, is less than a fixed value R.

$$n \geq \frac{4z_\alpha^2}{R^2} - \nu_0$$

16 Frequentist inference

16.1. Gaussian process: Symmetric $(1-2\alpha)\cdot 100\%$ confidence intervals

for gaussian processes from a normal distribution $\phi_{(\mu,\sigma)}$ from n observations with observables n, Σ_x and SS_x, and derived values $\nu = n-1$, $\bar{x} = \Sigma_x/n$ and $s_x^2 = SS_x/(n-1)$.

16.1.1 For μ, with known $\sigma = \sigma_0$: $\widehat{I}^{\mu}_{2\alpha} = \bar{x} \pm z_\alpha \cdot \sigma_0 \cdot \sqrt{\frac{1}{n}}$

16.1.2 For μ, with unknown σ: $\widehat{I}^{\mu}_{2\alpha} = \bar{x} \pm t_{\nu\,,\,\alpha} \cdot s_x \cdot \sqrt{\frac{1}{n}}$

16.1.3 For $\tau = \frac{1}{\sigma^2}$: $\widehat{I}^{\tau}_{2\alpha} = \left(\Gamma^{-1}_{\left(\frac{\nu}{2}\,,\,\frac{SS_x}{2}\right)}(\alpha)\ ,\ \Gamma^{-1}_{\left(\frac{\nu}{2}\,,\,\frac{SS_x}{2}\right)}(1-\alpha)\right)$

16.2. Gaussian process: Symmetric $(1-2\alpha)\cdot 100\%$ predictive intervals

for gaussian processes from a normal distribution $\phi_{(\mu,\sigma)}$ with n observations with observables n, Σ_x og SS_x, and derived values $\nu = n-1$, $\bar{x} = \Sigma_x/n$ and $s_x^2 = SS_x/(n-1)$.

16.2.1 For x_{n+1}, with known $\sigma = \sigma_0$: $\widehat{I}^{\mu}_{2\alpha} = \bar{x} \pm z_\alpha \cdot \sigma_0 \cdot \sqrt{1+\dfrac{1}{n}}$

16.2.2 For x_{n+1}, with unknown σ: $\widehat{I}^{\mu}_{2\alpha} = \bar{x} \pm t_{\nu\,,\,\alpha} \cdot s_x \cdot \sqrt{1+\dfrac{1}{n}}$

16.3. Bernoulli process: Symmetric $(1-2\alpha)\cdot 100\%$ confidence intervals

for a Bernoulli process with n tries and k positives, and derived value $\hat{\pi} = k/n$.

16.3.1 For the parameter π, an approximate interval: $\widehat{I}^{\pi}_{2\alpha} = \hat{\pi} \pm z_\alpha \cdot \sqrt{\dfrac{\hat{\pi}(1-\hat{\pi})}{n}}$

16.4. Hypothesis testing

with significance α is decided in a frequentist way like this: If the test is one-sided, and we have a p value calculated from one of the formulas below, our verdict is:

- If $p < \alpha$, we *reject* H_0 with significance α
- If $p \geq \alpha$, *don't* reject H_0 with significance α

If the test is two-sided, we reject H_0 with significance α *iff* we reject H_0 in one of the one-sided tests with significance $\frac{\alpha}{2}$.

16.5. Gaussian: Hypothesis test of μ

relative to a reference value μ_0, with significance α, for a Gaussian process from a normal distribution $\phi_{(\mu,\sigma)}$ with n measurements with calculated values $\nu = n-1$, $\bar{x}$ and s_x. The Null hypothesis H_0: $\mu = \mu_0$. For a right-sided test, the alternative hypothesis is H_1: $\mu > \mu_0$, whereas for a left-sided test, H_1: $\mu < \mu_0$.

16.5.1 Method if $\sigma = \sigma_0$ is known: Let $w = \dfrac{\bar{x} - \mu_0}{\sigma_0/\sqrt{n}}$.

For a left-sided test, let $p = \Phi(w)$, and for a right-sided test, let $p = \Phi(-w)$. Use 16.4 for the verdict.

16.5.2 Method if σ is unknown: Let $w = \dfrac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$

For a left-sided test, let $p = T_\nu(w)$, and for a right-sided test, let $p = T_\nu(-w)$. Use 16.4 for the verdict.

16.6. Gaussian process: Hypothesis test of σ

relative to a reference value σ_0, with significance α, for a gaussian process from a normal distribution $\phi_{(\mu,\sigma)}$ with n measurement with observables n, Σ_x and SS_x, og derived value $\nu = n - 1$. The Null hypothesis H_0: $\sigma = \sigma_0$. For a right-sided test, the alternative hypothesis is that H_1: $\sigma > \sigma_0$, whereas for a left-sided test, H_1: $\sigma < \sigma_0$.

16.6.1 Exact calculation: Let $\tau_0 = \frac{1}{\sigma_0^2}$, and let

$$p = \begin{cases} \Gamma_{\left(\frac{\nu}{2}, \frac{SS_x}{2}\right)}(\tau_0) & \text{(left-sided test)} \\ 1 - \Gamma_{\left(\frac{\nu}{2}, \frac{SS_x}{2}\right)}(\tau_0) & \text{(right-sided test)} \end{cases}$$

Use 16.4 for the verdict.

16.7. Bernoulli process: Hypothesis test of parameter π

relative to a reference value π_0, with significance α, for a Bernoulli process with n measurements with k positives, and derived value $\hat{\pi} = k/n$. The Null hypothesis is H_0: $\pi = \pi_0$. For a right-sided test, the alternative hypothesis is H_1: $\pi > \pi_0$, whereas for a left-sided test, H_1: $\pi < \pi_0$.

16.7.1 Method: Use the following value for p:

$$p = \begin{cases} \sum_{m=0}^{k} \binom{n}{m} \pi_0^m (1 - \pi_0)^{n-m} & \text{(left-sided test)} \\ \sum_{m=k}^{n} \binom{n}{m} \pi_0^m (1 - \pi_0)^{n-m} & \text{(right-sided test)} \end{cases}$$

Use 16.4 for the verdict.

16.7.2 Quick, approximate method: Let $w = \dfrac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$

For a left-sided test, let $p = \Phi(w)$, and for a right-sided test, let $p = \Phi(-w)$. Use 16.4 for the verdict.

17 Inference for the regression line

17.1. Matrix regression

17.1.1 Design matrix and response vector:
The starting point is the measurements (data) $\{(x_i, y_i)\}_{i=1}^{n}$, where x_i is the control variable. *The design matrix* X and *response vector* $\vec{y}$ are defined by:

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

17.1.2 The regression line: $y = \alpha + \beta x$ where the coefficients are given by

$$\vec{\beta} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \left(X^T X\right)^{-1} \cdot X^T \vec{y}$$

17.1.3 Centered form: Centered form for x data is $x_k^* = x_k - \bar{x}$. In the design matrix we then exchange x_k for $x_k^* = x_k - \overline{x}$, and the result is $\vec{\beta}_* = \begin{bmatrix} \alpha_* \\ \beta \end{bmatrix}$, with regression line $y = \alpha_* + \beta x^* = \alpha_* + \beta(x - \bar{x})$

17.1.4 The matrix $X^T X$ has contains the following useful information:

General: $X^T X = \begin{bmatrix} n & \Sigma_x \\ \Sigma_x & \Sigma_{x^2} \end{bmatrix}$ and in centered form, $X^T X = \begin{bmatrix} n & 0 \\ 0 & SS_x \end{bmatrix}$

17.1.5 Total error for the regression line: $SS_e = \vec{y}^T \vec{y} - \vec{\beta}^T \cdot (X^T X) \cdot \vec{\beta}$
$= \vec{y}^T \vec{y} - \vec{\beta}^T \cdot (X^T \vec{y})$

17.1.6 Squared standard error for the regression line: $s_e^2 = \frac{SS_e}{n-2}$

17.2. Bayes' theorem for linear regression

with n pairs $\{(x_i, y_i)\}_{i=1}^{n}$:

17.2.1 Informative prior hyperparameters for σ: Let σ_0 be your best estimate for σ, and n_0 be the equivalent number of measurements. Let $\nu_0 = n_0 - 2$ and $SS_0 = \sigma_0^2 \cdot \max(0, \nu_0)$.

17.2.2 Neutral prior hyperparameters for σ: $\nu_0 = -2$ and $SS_0 = 0$.

17.2.3 Updating: $P_1 \models \bar{x} = \dfrac{\Sigma_x}{n}$

$$\vec{\beta} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = (X^T X)^{-1} X^T \vec{y}$$

$$\nu_1 = \nu_0 + n$$

$$SS_1 = SS_0 + SS_e$$

17.2.4 Posterior values:

$$
\begin{aligned}
\tau &\sim \Gamma_{\left(\frac{\nu_1}{2}, \frac{SS_1}{2}\right)} \\
s_1^2 &= \frac{SS_1}{\nu_1} \\
b &\sim t_{\left(\beta\,,\, s_1\cdot\sqrt{\frac{1}{SS_x}}\,,\, \nu_1\right)} \\
y(x) &\sim t_{\left(\alpha_*+\beta(x-\bar{x})\,,\, s_1\cdot\sqrt{\frac{1}{n}+\frac{1}{SS_x}(x-\bar{x})^2}\,,\, \nu_1\right)} \\
Y_+(x) &\sim t_{\left(\alpha_*+\beta(x_+-\bar{x})\,,\, s_1\cdot\sqrt{1+\frac{1}{n}+\frac{1}{SS_x}(x-\bar{x})^2}\,,\, \nu_1\right)}
\end{aligned}
$$

Note that a is just $y(0)$, and therefore has the same distribution, whereas a_* follows $y(\bar{x})$.

17.2.5 $100(1-2\theta)\%$ (bayesian) credible and predictive intervals:

$$I_{2\theta}(x) = \alpha + \beta x \pm t_{\nu_1\,,\,\theta} \cdot s_1 \cdot \sqrt{\frac{1}{n} + \frac{1}{SS_x}(x-\bar{x})^2}$$

$$I_{2\theta}^{+}(x) = \alpha + \beta x \pm t_{\nu_1\,,\,\theta} \cdot s_1 \cdot \sqrt{1 + \frac{1}{n} + \frac{1}{SS_x}(x-\bar{x})^2}$$

17.3. Frequentist linear regression

employs the same basic formulas as bayesian linear regression, but for the interval estimates, use $s_e = \sqrt{\frac{SS_e}{n-2}}$ instead of s_1, and $n-2$ instead of ν_1.

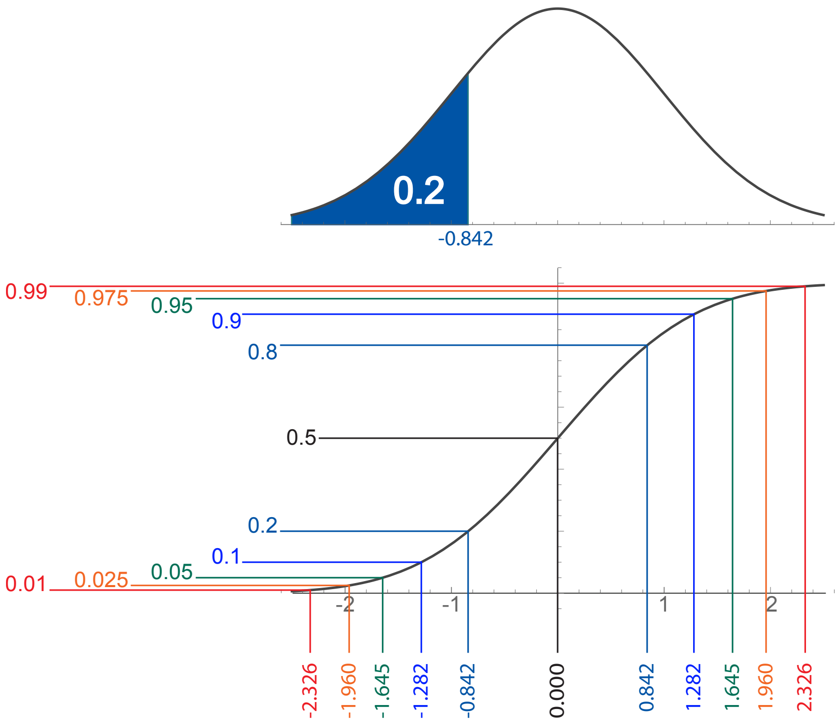
17.3.1 $100(1-2\theta)\%$ (frequentist) confidence and predictive intervals:

$$\widehat{I}_{2\theta}(x) = \alpha + \beta x \pm t_{n-2\,,\,\theta} \cdot s_e \cdot \sqrt{\frac{1}{n} + \frac{1}{SS_x}(x-\bar{x})^2}$$

$$\widehat{I}_{2\theta}^{+}(x) = \alpha + \beta x \pm t_{n-2\,,\,\theta} \cdot s_e \cdot \sqrt{1 + \frac{1}{n} + \frac{1}{SS_x}(x-\bar{x})^2}$$

18 Tables

18.1. $z_p = \Phi^{-1}_{(0,1)}(p)$ Percentiles for Normal distribution



p	z_p
0.00000	$-\infty$
0.0001	-3.719
0.00025	-3.481
0.0005	-3.290
0.001	-3.090
0.0025	-2.807
0.005	-2.576
0.01	-2.326
0.015	-2.170
0.02	-2.054
0.025	-1.960
0.03	-1.881
0.035	-1.812
0.04	-1.751
0.045	-1.695
0.05	-1.645
0.06	-1.555

p	z_p
0.07	-1.476
0.08	-1.405
0.09	-1.341
0.10	-1.282
0.15	-1.036
0.20	-0.842
0.30	-0.524
0.40	-0.253
0.50	**0**
0.60	0.253
0.70	0.524
0.80	0.842
0.85	1.036
0.90	1.282
0.91	1.341
0.92	1.405
0.93	1.476

p	z_p
0.94	1.555
0.95	1.645
0.955	1.695
0.96	1.751
0.965	1.812
0.97	1.881
0.975	1.960
0.98	2.054
0.985	2.170
0.99	2.326
0.995	2.576
0.9975	2.807
0.999	3.090
0.9995	3.290
0.99975	3.481
0.9999	3.719
1.00000	∞

Values for the right tail have the opposite sign, since z is anti symmetrical around $p = 0.5$:

$$z_{1-p} = -z_p$$

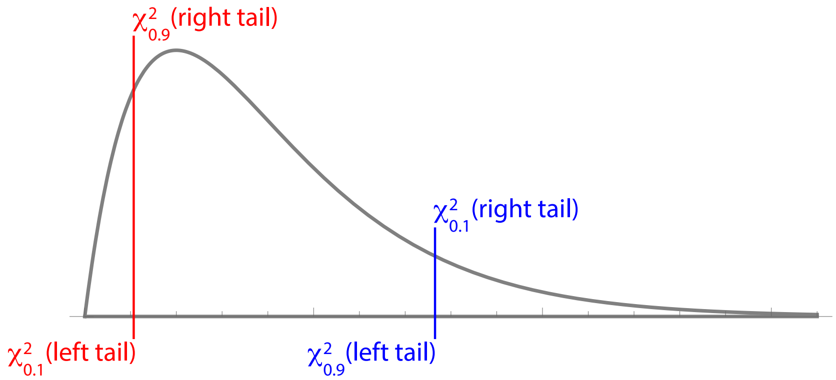
18.2. Percentiles for Student's t with ν degrees of freedom

Values for $-T^{-1}_{(0,1,\nu)}(p) = -t_{\nu,p} = t_{\nu,1-p}$. For $\nu > 30$, use the normal approximation ($\nu = \infty$).

ν \ p	0.1	0.075	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005	0.00025	0.0001
1	3.0777	4.1653	6.3138	12.706	31.821	63.657	127.32	318.31	636.62	1273.2	3183.1
2	1.8856	2.2819	2.9200	4.3027	6.9646	9.9248	14.089	22.327	31.599	44.705	70.700
3	1.6377	1.9243	2.3534	3.1824	4.5407	5.8409	7.4533	10.215	12.924	16.326	22.204
4	1.5332	1.7782	2.1318	2.7764	3.7469	4.6041	5.5976	7.1732	8.6103	10.306	13.034
5	1.4759	1.6994	2.0150	2.5706	3.3649	4.0321	4.7733	5.8934	6.8688	7.9757	9.6776
6	1.4398	1.6502	1.9432	2.4469	3.1427	3.7074	4.3168	5.2076	5.9588	6.7883	8.0248
7	1.4149	1.6166	1.8946	2.3646	2.9980	3.4995	4.0293	4.7853	5.4079	6.0818	7.0634
8	1.3968	1.5922	1.8595	2.3060	2.8965	3.3554	3.8325	4.5008	5.0413	5.6174	6.4420
9	1.3830	1.5737	1.8331	2.2622	2.8214	3.2498	3.6897	4.2968	4.7809	5.2907	6.0101
10	1.3722	1.5592	1.8125	2.2281	2.7638	3.1693	3.5814	4.1437	4.5869	5.0490	5.6938
11	1.3634	1.5476	1.7959	2.2010	2.7181	3.1058	3.4966	4.0247	4.4370	4.8633	5.4528
12	1.3562	1.5380	1.7823	2.1788	2.6810	3.0545	3.4284	3.9296	4.3178	4.7165	5.2633
13	1.3502	1.5299	1.7709	2.1604	2.6503	3.0123	3.3725	3.8520	4.2208	4.5975	5.1106
14	1.3450	1.5231	1.7613	2.1448	2.6245	2.9768	3.3257	3.7874	4.1405	4.4992	4.9850
15	1.3406	1.5172	1.7531	2.1314	2.6025	2.9467	3.2860	3.7328	4.0728	4.4166	4.8800
16	1.3368	1.5121	1.7459	2.1199	2.5835	2.9208	3.2520	3.6862	4.0150	4.3463	4.7909
17	1.3334	1.5077	1.7396	2.1098	2.5669	2.8982	3.2224	3.6458	3.9651	4.2858	4.7144
18	1.3304	1.5037	1.7341	2.1009	2.5524	2.8784	3.1966	3.6105	3.9216	4.2332	4.6480
19	1.3277	1.5002	1.7291	2.0930	2.5395	2.8609	3.1737	3.5794	3.8834	4.1869	4.5899
20	1.3253	1.4970	1.7247	2.0860	2.5280	2.8453	3.1534	3.5518	3.8495	4.1460	4.5385
21	1.3232	1.4942	1.7207	2.0796	2.5176	2.8314	3.1352	3.5272	3.8193	4.1096	4.4929
22	1.3212	1.4916	1.7171	2.0739	2.5083	2.8188	3.1188	3.5050	3.7921	4.0769	4.4520
23	1.3195	1.4893	1.7139	2.0687	2.4999	2.8073	3.1040	3.4850	3.7676	4.0474	4.4152
24	1.3178	1.4871	1.7109	2.0639	2.4922	2.7969	3.0905	3.4668	3.7454	4.0207	4.3819
25	1.3163	1.4852	1.7081	2.0595	2.4851	2.7874	3.0782	3.4502	3.7251	3.9964	4.3517
26	1.3150	1.4834	1.7056	2.0555	2.4786	2.7787	3.0669	3.4350	3.7066	3.9742	4.3240
27	1.3137	1.4817	1.7033	2.0518	2.4727	2.7707	3.0565	3.4210	3.6896	3.9538	4.2987
28	1.3125	1.4801	1.7011	2.0484	2.4671	2.7633	3.0469	3.4082	3.6739	3.9351	4.2754
29	1.3114	1.4787	1.6991	2.0452	2.4620	2.7564	3.0380	3.3962	3.6594	3.9177	4.2539
30	1.3104	1.4774	1.6973	2.0423	2.4573	2.7500	3.0298	3.3852	3.6460	3.9016	4.2340
∞	1.2816	1.4395	1.6449	1.9600	2.3263	2.5758	2.8070	3.0902	3.2905	3.4808	3.7190

18.3. Percentiles for the χ^2 distribution with ν degrees of freedom (left tail)

The table shows χ^2_p for the left tail. You find the value for p for the right tail as the value of χ^2_{1-p}. Notice that due to the unsymmetrical nature of χ^2, we can't make use of any symmetries.



ν \ p	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.07	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892

18.4. The Γ *function*

	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	1	9.51351	4.59084	2.99157	2.21816	1.77245	1.48919	1.29806	1.16423	1.06863
1	1	0.951351	0.918169	0.897471	0.887264	0.886227	0.893515	0.908639	0.931384	0.961766
2	1	1.04649	1.1018	1.16671	1.24217	1.32934	1.42962	1.54469	1.67649	1.82736
3	2	2.19762	2.42397	2.68344	2.98121	3.32335	3.71702	4.17065	4.69417	5.29933
4	6	6.81262	7.75669	8.85534	10.1361	11.6317	13.3813	15.4314	17.8379	20.6674
5	24	27.9318	32.5781	38.078	44.5988	52.3428	61.5539	72.5276	85.6217	101.27
6	120	142.452	169.406	201.813	240.834	287.885	344.702	413.408	496.606	597.494
7	720	868.957	1050.32	1271.42	1541.34	1871.25	2275.03	2769.83	3376.92	4122.71
8	5040	6169.59	7562.29	9281.39	11405.9	14034.4	17290.2	21327.7	26340.	32569.4
9	40320	49973.7	62010.8	77035.6	95809.5	119292	148696	185551	231792	289868

	0	1	2	3	4	5	6	7	8	9
10	$3.63\cdot10^{5}$	$3.63\cdot10^{6}$	$3.99\cdot10^{7}$	$4.79\cdot10^{8}$	$6.23\cdot10^{9}$	$8.72\cdot10^{10}$	$1.31\cdot10^{12}$	$2.09\cdot10^{13}$	$3.56\cdot10^{14}$	$6.40\cdot10^{15}$
20	$1.22\cdot10^{17}$	$2.43\cdot10^{18}$	$5.11\cdot10^{19}$	$1.12\cdot10^{21}$	$2.59\cdot10^{22}$	$6.2\cdot10^{23}$	$1.55\cdot10^{25}$	$4.03\cdot10^{26}$	$1.09\cdot10^{28}$	$3.05\cdot10^{29}$
30	$8.84\cdot10^{30}$	$2.65\cdot10^{32}$	$8.22\cdot10^{33}$	$2.63\cdot10^{35}$	$8.68\cdot10^{36}$	$2.95\cdot10^{38}$	$1.03\cdot10^{40}$	$3.72\cdot10^{41}$	$1.38\cdot10^{43}$	$5.23\cdot10^{44}$
40	$2.04\cdot10^{46}$	$8.16\cdot10^{47}$	$3.35\cdot10^{49}$	$1.41\cdot10^{51}$	$6.04\cdot10^{52}$	$2.66\cdot10^{54}$	$1.2\cdot10^{56}$	$5.5\cdot10^{57}$	$2.59\cdot10^{59}$	$1.24\cdot10^{61}$
50	$6.08\cdot10^{62}$	$3.04\cdot10^{64}$	$1.55\cdot10^{66}$	$8.07\cdot10^{67}$	$4.27\cdot10^{69}$	$2.31\cdot10^{71}$	$1.27\cdot10^{73}$	$7.11\cdot10^{74}$	$4.05\cdot10^{76}$	$2.35\cdot10^{78}$
60	$1.39\cdot10^{80}$	$8.32\cdot10^{81}$	$5.08\cdot10^{83}$	$3.15\cdot10^{85}$	$1.98\cdot10^{87}$	$1.27\cdot10^{89}$	$8.25\cdot10^{90}$	$5.44\cdot10^{92}$	$3.65\cdot10^{94}$	$2.48\cdot10^{96}$
70	$1.71\cdot10^{98}$	$1.2\cdot10^{100}$	$8.5\cdot10^{101}$	$6.1\cdot10^{103}$	$4.5\cdot10^{105}$	$3.3\cdot10^{107}$	$2.5\cdot10^{109}$	$1.9\cdot10^{111}$	$1.5\cdot10^{113}$	$1.1\cdot10^{115}$
80	$8.9\cdot10^{116}$	$7.2\cdot10^{118}$	$5.8\cdot10^{120}$	$4.8\cdot10^{122}$	$3.9\cdot10^{124}$	$3.3\cdot10^{126}$	$2.8\cdot10^{128}$	$2.4\cdot10^{130}$	$2.1\cdot10^{132}$	$1.9\cdot10^{134}$
90	$1.7\cdot10^{136}$	$1.5\cdot10^{138}$	$1.4\cdot10^{140}$	$1.2\cdot10^{142}$	$1.2\cdot10^{144}$	$1.1\cdot10^{146}$	$1.\cdot10^{148}$	$9.9\cdot10^{149}$	$9.6\cdot10^{151}$	$9.4\cdot10^{153}$
100	$9.3\cdot10^{155}$	$9.3\cdot10^{157}$	$9.4\cdot10^{159}$	$9.6\cdot10^{161}$	$9.9\cdot10^{163}$	$1.\cdot10^{166}$	$1.1\cdot10^{168}$	$1.1\cdot10^{170}$	$1.2\cdot10^{172}$	$1.3\cdot10^{174}$
110	$1.4\cdot10^{176}$	$1.6\cdot10^{178}$	$1.8\cdot10^{180}$	$2.\cdot10^{182}$	$2.2\cdot10^{184}$	$2.5\cdot10^{186}$	$2.9\cdot10^{188}$	$3.4\cdot10^{190}$	$4.\cdot10^{192}$	$4.7\cdot10^{194}$
120	$5.6\cdot10^{196}$	$6.7\cdot10^{198}$	$8.1\cdot10^{200}$	$9.9\cdot10^{202}$	$1.2\cdot10^{205}$	$1.5\cdot10^{207}$	$1.9\cdot10^{209}$	$2.4\cdot10^{211}$	$3.\cdot10^{213}$	$3.9\cdot10^{215}$
130	$5.\cdot10^{217}$	$6.5\cdot10^{219}$	$8.5\cdot10^{221}$	$1.1\cdot10^{224}$	$1.5\cdot10^{226}$	$2.\cdot10^{228}$	$2.7\cdot10^{230}$	$3.7\cdot10^{232}$	$5.\cdot10^{234}$	$6.9\cdot10^{236}$
140	$9.6\cdot10^{238}$	$1.3\cdot10^{241}$	$1.9\cdot10^{243}$	$2.7\cdot10^{245}$	$3.9\cdot10^{247}$	$5.6\cdot10^{249}$	$8.\cdot10^{251}$	$1.2\cdot10^{254}$	$1.7\cdot10^{256}$	$2.6\cdot10^{258}$

For $n \geq 10$, *Stirling's approximation* to Γ (1.3.4) gives a value less than 1% off the true value:

$$\Gamma(x) \approx \left(\frac{x}{e}\right)^x \cdot \sqrt{\frac{2\pi}{x}}$$

For positive integers n, use that Γ is a generalized faculty to get an exact answer:

$\Gamma(n) = (n-1)!$ and $\Gamma(n+\frac{1}{2}) = \frac{(2n)!}{4^n n!}\sqrt{\pi}$