Discrete probability distributions

X	X Counts	p(x)	Values of X	E(x)	V(x)
Discrete uniform	Outcomes that are equally likely (finite)	$\frac{1}{b-a+1}$	$a \le x \le b$	$\frac{b+a}{2}$	$\frac{(b-a+2)(b-a)}{12}$
Binomial	Number of sucesses in n fixed trials	$\binom{n}{x} p^x (1-p)^{n-x}$	$x = 0,1,\ldots,n$	np	$np(1-p)$
Poisson	Number of arrivals in a fixed time period	$\frac{e^{-\lambda}\lambda^x}{x!}$	$x = 0,1,2,\ldots$	λ	λ
Geometric	Number of trials up through 1st success	$(1-p)^{x-1}p$	$x = 1,2,3,\ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial	Number of trials up through kth success	$\binom{x-1}{k-1}(1-p)^{x-k}p^k$	$x = k, k+1,\ldots$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$
Hyper - geometric	Number of marked individuals in sample taken without replacement	$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$\max(0, M+n-N) \le x \le \min(M,n)$	$n * \frac{M}{N}$	$\frac{nM(N-M)(N-n)}{N^2(N-1)}$

Continuous probability distributions

- Normal distribution: $N(\mu, \sigma^2)$
  Density function: $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Log normal distribution: $LogN(\mu, \sigma^2)$
  Y is lognormal, $Y \sim LogN(\mu, \sigma^2)$, if $Y = e^X$ where $X \sim N(\mu, \sigma^2)$.
  All formulas below are in terms of the parameters of the underlying normal distribution.

  Density function: $\frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$
  Mean: $e^{\left(\mu+\frac{\sigma^2}{2}\right)}$
  Variance: $\left(e^{\sigma^2}-1\right)e^{(2\mu+\sigma^2)}$
  Median: e^{μ}

Wilcoxon Rank Sum for non-matched samples

Definition:

Consider observed values $v_1 \leq v_2 \leq \ldots \leq v_N$, coming from two classes, C_1 and C_2. In this representation $1 \leq i \leq N$ is the rank of the value v_i. The class associated to each observation is given by a vector of labels:
LABELS = $L_1, \ldots, L_N$, matching the above ranking. That is: each L_i is either 1 or 2, depending on whether the observed vi came from C_1 or from C_2.
We will compute the sum of ranks for C_1 to get the WRS statistic T:

$$T = \sum_{i=1}^{N} i \cdot \mathbf{1}[Li = 1]$$

Where **1**[BOOL] is an indicator of the Boolean BOOL.
Let B = size of C_1.
Under the null model where LABELS is uniformly distributed amongst all possible $\binom{N}{B}$ configurations we have

$$E(T) = \frac{B(N+1)}{2}$$

Normal approximations

- For the Wilcoxon Rank Sum statistic, T.

  Let $\mu_T = \frac{B(N+1)}{2}$ and $\sigma_T = \sqrt{\frac{B(N-B)(N+1)}{12}}$.

  Then $Z(T) = \frac{T-\mu_T}{\sigma_T}$ has an approximate standard normal distribution.
- For Spearman correlations (Fisher Transform).

  For -1 < r < 1 let $F(r) = \frac{1}{2} \ln \frac{1+r}{1-r}$.

  Also denote the Spearman correlation of two independent uniformly drawn permutations over n numbers by ρ_n.

  Then $Z = \sqrt{\frac{n-3}{1.06}}\, F(\rho_n)$ has a distribution that is well approximated by a standard normal distribution

Markov

- The Markov transition matrix is the matrix T such that:
$$T(i,j) = P(X_1 = j | X_0 = i)$$
- $\pi = (\pi_1, \dots, \pi_k)$ is a stationary distribution for a Markov chain over the states $1, \dots, k$ if:
$$\pi \cdot T = \pi$$

Other

- Conversion to standard normal:
If $X \sim N(\mu, \sigma)$ then $Z = \frac{X - \mu}{\sigma}$ has a standard normal distribution
- Convolution:
Let X and Y be independent random variables taking values in the integers.
Let Z = X+Y.
Then:
$$\forall z \quad P(Z = z) = \sum_{i=-\infty}^{\infty} P(X = i)P(Y = z - i)$$
- Entropy.
Shannon's entropy for a random variable X that takes values in a finite set indexed by $i = 1 \dots n$, (possibly n is infinity) with probabilities p_i:
$$H(X) = -\sum_{i=1}^{n} p_i \log(p_i)$$
- Pearson correlation
Let $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ be the observed data.
$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$
Where $\bar{x} = \frac{1}{n}\sum_1^n x_i$ and $\bar{y} = \frac{1}{n}\sum_1^n y_i$

- Rank correlations
  Let $(x_1,\ y_1), (x_2,\ y_2), \dots, (x_n,\ y_n)$ be the observed data. Assume that all values of (x_i) and (y_i) are unique.
  - Spearman, when all ranks are distinct:

$$\rho(x, y) = 1 - \frac{6 \sum d_i^2}{n(n2 - 1)}$$

    where $d_i = rank(x_i) - rank(y_i)$

  - Kendall
    A pair of observations $(x_i,\ y_i)$ and $(x_j,\ y_j)$ is said to be *concordant* if the ranks for both elements agree: that is, if both
    $x_i < x_j$ and $y_i > y_j$ OR if both $x_i < x_j$ and $y_i < y_j$.
    It is said to be *discordant*, if $x_i > x_j$ and $y_i < y_j$ OR if $x_i < x_j$ and $y_i > y_j$.
    If $x_i = x_j$ or $y_i = y_j$, the pair is neither concordant nor discordant (and we assumed, for simplicity, that this doesn't happen).
    Let C and D be the number of concordant and discordant pairs, respectively. The Kendall τ coefficient is defined as:

$$\tau = \frac{C - D}{\binom{n}{2}}$$

Integration by Parts

$$\int u(x)v'(x)dx = u(x)v(x) - \int v(x)u'(x)\,dx$$

Confidence Intervals

- Confidence interval for empirical proportion $\hat{p}$:

  $p \in [\hat{p} \ \pm \gamma\hat{\sigma}]$ with probability $(1-\alpha)$,

  where $\gamma = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$, and $\hat{\sigma} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

- For confidence interval with difference in proportions:

$$\Delta \in \ (\hat{p}_1 \ - \hat{p}_2) \ \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \ \left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)^{\frac{1}{2}}$$

  with probability $(1-\alpha)$.

Credibility Intervals

Prior: $\pi(\theta)$

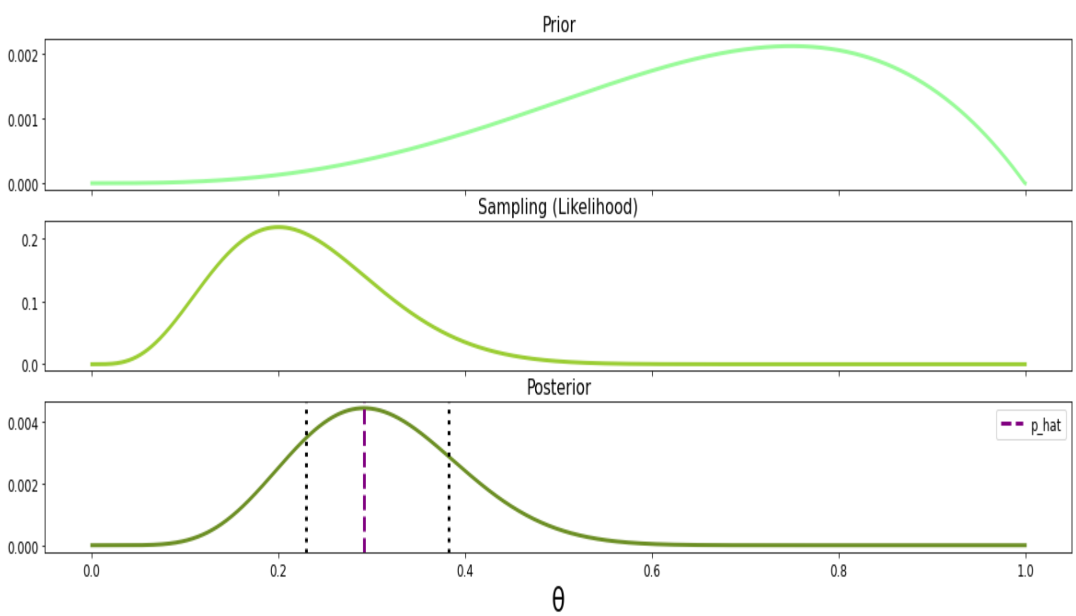
Likelihood – having observed the data we compute

$$L(\theta) = Prob(D;\ \theta)$$

In the Bayes approach we compute a posterior:

$$\psi(\theta) = \pi(\theta)L(\theta)$$

We return $\hat{\theta} = \mathrm{argmax}\,\psi(\theta)$ and compute credibility intervals using ψ



Z-Score Table

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
+0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
+0.1	.53983	.54380	.54776	.55172	.55567	.55966	.56360	.56749	.57142	.57535
+0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
+0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
+0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
+0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
+0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
+0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
+0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
+0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
+1	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
+1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
+1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
+1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
+1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
+1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
+1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
+1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
+1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
+1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
+2	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
+2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
+2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
+2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
+2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
+2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
+2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
+2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
+2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
+2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
+3	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
+3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
+3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
+3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
+3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
+3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
+3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
+3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
+3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
+3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997
+4	.99997	.99997	.99997	.99997	.99997	.99997	.99998	.99998	.99998	.99998