

MISC	PROB BASICS	DISCRETE DISTROS	CONDITIONAL VARS	EXPECTATIONS										
Log $\ln(mn) = \ln(m) + \ln(n)$ $\ln\left(\frac{m}{n}\right) = \ln(m) - \ln(n)$ $\ln(m^r) = r \ln(m)$ Exponent $(ab)^x = a^x b^x$ $(a^x)^y = a^{xy}$ $a^x a^y = a^{x+y}$ Summation $\sum_{i=1}^n ar^{i-1} = a \frac{1-r^n}{1-r}$ Integrals $\int \frac{1}{x} dx = \ln x $ $\int e^{ax} dx = \frac{1}{a} e^{ax}$ $\int a^x dx = \frac{a^x}{\ln(a)}$ $\int \ln(x) dx = x \ln(x) - x$ $\int \cos(x) dx = \sin(x)$ $\int \sin(x) dx = -\cos(x)$ Derivatives $(e^x)' = e^x$ $(\ln(x))' = \frac{1}{x}$ $\sin(x) = \cos(x)$ $\cos(x) = -\sin(x)$ $(fg)' = fg' + f'g$ $\frac{1}{f} = -\frac{f'}{f^2}$ $(f(g(x)))' = f'(g(x))g'(x)$ $\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}$ Median Middle number in sorted. If discrete distro, check up to where we have $p < 0.5$ and then $p > 0.5$, the number we have to add to cross threshold is median (see here)	Properties $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ $P(A \cup B) \leq P(A) + P(B)$ Conditional $P(A B) = \frac{P(A \cap B)}{P(B)}$ $P(A \cap B) = P(B)P(A B)$ △ Total Prob Theorem △ $P(B) = P(A_1 \cap B) + \dots + P(A_n \cap B)$ $= P(A_1)P(B A_1) + \dots$ Bayes $P(A B) = \frac{P(A)P(B A)}{P(B)}$	Bernouilli $P(X = 1) = p$ $E[X] = p$ $Var(X) = p(1-p)$ Uniform DISCRETE $p_X(x) = \frac{1}{b-a+1}$ $F_X(k) = \frac{\lfloor k \rfloor - a + 1}{n}$ $E[X] = \frac{a+b}{2}$ $Var(X) = \frac{(b-a+1)^2 - 1}{12}$ Binomial k successes in n trials $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ $E[X] = np$ $Var(X) = np(1-p)$ Geometric number of trials until success $p_X(k) = (1-p)^{k-1} p$ $F_X(k) = 1 - (1-p)^k$ $E[X] = \frac{1}{p}$ $Var(X) = \frac{1-p}{p^2}$ Poisson how many occurrences k in τ given rate λ $P(k, \tau) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}$ $I = \frac{1}{\lambda}$ $E[N_\tau] = \lambda\tau$ $Var(N_\tau) = \lambda\tau$	same for PDF $p_{X Y}(x y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$ Multiplication Rule $p_{X,Y,Z}(x, y, z) = p_X(x)p_{Y X}(y x)p_{Z XY}(z x, y)$ $p_{X,Y Z}(x, y z) = \frac{p_{X,Y,Z}(x, y, z)}{p_Z(z)}$	Expected Value $E[g(x)] = \sum_x g(x)p_X(x)$ $E[g(x)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$ Linearity of Expectations $E[aX + b] = aE[X] + b$ $E[X + Y] = E[X] + E[Y]$ Total Expectation Th. $E[X] = \sum_y p_Y(y)E[X Y = y]$ $E[X] = \int_{-\infty}^{\infty} f_Y(y)E[X Y = y]dy$ $E[X] = \sum_i P(A_i)E[X A_i]$ Cond. Expectation $E[g(x) Y = y] = \sum_x g(x)p_{X Y}(x y)$ Iterated Expectation $E[E[X Y]] = E[X] \quad (\text{ex. } \underline{\text{ex.}})$										
CONT. DISTROS $P(a \leq x \leq b) = \int_a^b f_X(x)dx$ Disjoint $P(1 \leq x \leq 3 \text{ or } 4 \leq x \leq 5) = P(1 \leq x \leq 3) + P(4 \leq x \leq 5)$ Properties $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f_X(x)dx = 1$ CDF Properties <ul style="list-style-type: none"> • $\rightarrow_{x \rightarrow \infty} 1$ and $\rightarrow_{x \rightarrow -\infty} 0$ • increasing/monotonic • right-continuous Uniform CONT $f_X(x) = \frac{1}{b-a}$ $\hat{b}^{\text{MLE}} = \max(X_i)$ $F_X(x) = \frac{x-a}{b-a}$ $E[X] = \frac{a+b}{2}$ $Var(X) = \frac{(b-a)^2}{12}$ Exponential time to wait for something $f_X(x) = \lambda e^{-\lambda x}$ $P(X \geq a) = \int_a^{\infty} \lambda e^{-\lambda x} dx = e^{-\lambda a}$ $F_X(x) = 1 - e^{-\lambda x}$ $E[X] = \frac{1}{\lambda}$ $I = \frac{1}{\lambda^2}$ $E[X^2] = \frac{2}{\lambda^2}$ COUNTING n choose k nb of combinations (any order) $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ permutations nb of ways of ordering n elements (order matters) $n!$ subsets of n elements 2^n partitions n objects into r groups $n!$ $n_1! n_2! \dots n_r!$	$\int_{(x,y) \in B} f_{X,Y}(x,y) dxdy$ Marginals / Total Probability $p_X(x) = \sum_y p_{X,Y}(x,y) = \sum_y p_Y(y)p_{X Y}(x y)$ $f_X(x) = \int f_Y(y)f_{X Y}(x y)dy$ △ ranges: what values can Y take when X = x? $= \int f_{X,Y}(x,y)dy$ Expected Value Rule $E[g(X, Y)] = \sum_x \sum_y g(x, y)p_{X,Y}(x, y)$ $E[g(X, Y)] = \int E[g(x, y) Y = y]f_Y(y)dy$ CDF $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$ $= \int_y^y \int_{-\infty}^x f_{X,Y}(s, t)dsdt$	INDEPENDENCE If Indie $E[XY] = E[X]E[Y]$ $Var(X + Y) = Var(X) + Var(Y)$ $p_{X,Y Z}(x, y z) = p_{X Z}(x z)p_{Y Z}(y z)$ $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ $f_{X Y}(x y) = f_X(x)$ $Cov(X, Y) = 0$												
NORMALS $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ $E[X] = \mu$ $I(\mu, \sigma^2) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 1/(2\sigma^4) \end{pmatrix}$ $Var(X) = \sigma^2 \quad Var(X^2) = 2\sigma^4$ Linear Functions $Y = aX + b$ with $X \sim N(\mu, \sigma^2)$ $\hat{\mu}^{\text{MLE}} = \bar{X}_n$ $\hat{\sigma}^2 = S_n$ $Y = N(a\mu + b, a^2\sigma^2)$ (sample var)	VARIANCE $Var(X) = E[(X - \mu)^2]$ and $\sigma = \sqrt{Var(X)}$ Properties $Var(aX + b) = a^2 Var(X)$ $Var(X) = E[X^2] - (E[X])^2$ Dependent Sum $Var(X_1 + \dots + X_n) = \sum_{i=1}^n Var(X_i) + \sum_{\{i,j\}: i \neq j} Cov(X_i, X_j)$ Indie Sum $Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$ nb: $Z \sim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2)$ if $Z = N_1 - N_2$ Tables $\Phi(-2) = P(Y \leq -2) = 1 - P(Y \leq 2) = 1 - \Phi(2)$ Standardising $X \sim N(\mu, \sigma^2)$ and $Y \sim N(0, 1)$ $Y = \frac{X - \mu}{\sigma}$ $X = \mu + \sigma Y$ Moments <table border="0" style="width:100%; text-align:center;"> <tr> <td>$1 \ \mu$</td> <td>0</td> </tr> <tr> <td>$2 \ \mu^2 + \sigma^2$</td> <td>σ^2</td> </tr> <tr> <td>$3 \ \mu^3 + 3\mu\sigma^2$</td> <td>0</td> </tr> <tr> <td>$4 \ \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$</td> <td>$3\sigma^4$</td> </tr> <tr> <td>$5 \ \mu^5 + 10\mu^3\sigma^2 + 15\mu\sigma^4$</td> <td>$0$</td> </tr> <tr> <td>$6 \ \mu^6 + 15\mu^4\sigma^2 + 45\mu^2\sigma^4 + 15\sigma^6$</td> <td>$15\sigma^6$</td> </tr> </table>	$1 \ \mu$	0	$2 \ \mu^2 + \sigma^2$	σ^2	$3 \ \mu^3 + 3\mu\sigma^2$	0	$4 \ \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$	$3\sigma^4$	$5 \ \mu^5 + 10\mu^3\sigma^2 + 15\mu\sigma^4$	0	$6 \ \mu^6 + 15\mu^4\sigma^2 + 45\mu^2\sigma^4 + 15\sigma^6$	$15\sigma^6$	LAW OF TOTAL VAR $Var(X) = E[Var(X Y)] + Var(E[X Y])$ Sample Variance $S_n = \frac{1}{n} \sum (X_i - \bar{X}_n)^2$ $E[S_n] = \frac{n-1}{n} \sigma^2$ Unbiased Sample Variance $\widetilde{S}_n = \frac{n}{n-1} S_n$ $E[\widetilde{S}_n] = \sigma^2$
$1 \ \mu$	0													
$2 \ \mu^2 + \sigma^2$	σ^2													
$3 \ \mu^3 + 3\mu\sigma^2$	0													
$4 \ \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$	$3\sigma^4$													
$5 \ \mu^5 + 10\mu^3\sigma^2 + 15\mu\sigma^4$	0													
$6 \ \mu^6 + 15\mu^4\sigma^2 + 45\mu^2\sigma^4 + 15\sigma^6$	$15\sigma^6$													
STATISTICAL MODEL $(E, (P_\theta)_{\theta \in \Theta})$ E : sample space (X_1, \dots) P : family of prob measures on E Θ : Param set well specified if $\theta^* \in \Theta$ △ sample space must not depend on parameter △ sample space must be the support for the distribution. i.e. $([0, \infty), \{N(\mu, \sigma^2)\})$ is not valid because the sample space for a N is all R	RANDOM NB OF RANDOM VARIABLES N: nb of stores visited X_i : money spent in store i $Y = \sum X_i$ $E[Y] = E[N]E[X]$ $Var(Y) = E[N]var(X) + (E[X])^2 var(N)$													

<p>DERIVED DISTROS</p> <p>PMF function of discrete RV</p> $p_Y(y) = P(g(x) = y) = \sum_{x:g(x)=y} p_X(x)$ <p>Linear Functions</p> $Y = aX + b$ $p_Y(y) = p_X\left(\frac{y-b}{a}\right)$ $f_Y(y) = \frac{1}{ a } f_X\left(\frac{y-b}{a}\right)$ <p>g is monotonic</p> $f_Y(y) = f_X(h(y)) \left \frac{dh}{dy}(y) \right \text{ where } h \text{ is inverse of } g$ <p>general case</p> <ol style="list-style-type: none"> 1) find CDF: $F_Y(y) = P(g(x) \leq y)$ 2) derive CDF for y to find PDF 	<p>BERNOULLI PROCESS</p> <p>requires indie, time homogen.</p> <p>Properties</p> $S = X_1 + \dots + X_n$ $P(S = K) = \binom{n}{k} p^k (1-p)^{n-k}$ $E[S] = np$ $\text{Var}(S) = np(1-p)$ <p>Time until 1st success</p> $T_1 = \min\{i: X_i = 1\}$ $P(T_1 = k) = (1-p)^{k-1}p$ <p>Time of kth arrival</p> $p_{Y_k}(t) = \binom{t-1}{k-1} p^k (1-p)^{t-k}$ $E[Y_k] = \frac{k}{p} \triangleq \text{memoryless (ex.)}$ $\text{Var}(Y_k) = \frac{k(1-p)}{p^2}$	<p>POISSON PROCESS</p> <p>indie, time homogen. seq of exp</p> <p>λ: arrival rate</p> $P(k, \tau) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}$ $I = \frac{1}{\lambda}$ $E[N_\tau] = \lambda\tau$ $\text{Var}(N_\tau) = \lambda\tau$ $\lambda = \frac{E[N_\tau]}{\tau}$ <p>Time of kth arrival / Erlang</p> $f_{Y_k}(y) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}$ $= \text{Erlang}(k)$ $= \text{Erlang}\left(\frac{k}{2}\right) + \text{Erlang}\left(\frac{k}{2}\right)$ <p>Sum</p> <p>\triangleq must be indie</p> <p>M: Poisson(μ) N: Poisson(v)</p> <p>M+N: Poisson($\mu+v$)</p> <p>Merging</p> <p>A: λ_A B: λ_B</p> $\lambda = \lambda_A + \lambda_B$ $P(k^{\text{th}} \text{ arrival is A}) = \frac{\lambda_A}{\lambda_A + \lambda_B}$ <p>P(k arrivals are A) is Binomial$\left(\frac{\lambda_A}{\lambda_A + \lambda_B}\right)$</p> <p>Splitting</p> <p>flip a coin with prob q</p> <p>A-Ber(qp)</p> <p>B-Ber($(1-q)p$)</p> <p>\triangleq these streams are not indie</p>	<p>COVARIANCE MATRIX AND MV STUFF</p> $\Sigma = \begin{pmatrix} \text{Cov}(X, X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \text{Cov}(Y, Y) \end{pmatrix}$ $= E[(X - E[X])(Y - E[Y])^T]$ <p>ESTIMATORS</p> <p>Asym. normal if</p> $\sqrt{n}(\widehat{\theta}_n - \theta) \rightarrow N(0, \sigma^2)$ <p>Consistency</p> $\widehat{\theta}_n \rightarrow \theta \text{ as } n \rightarrow \infty$ <p>Bias</p> $\text{bias}(\widehat{\theta}_n) = E[\widehat{\theta}_n] - \theta$ <p>Quadratic Risk</p> $R(\widehat{\theta}_n) = E[(\widehat{\theta}_n - \theta)^2]$ <p>Confidence Interval level $1 - \alpha$</p> <p>conf.int. can't depend on unknown</p> $P\left(\overline{X}_n - q\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n + q\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$										
<p>CONVOLUTIONS</p> $Z = X + Y$ $p_Z(z) = \sum_x p_X(x)p_Y(z-x)$ $f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$	<p>Merging</p> $Z_t = g(X_t, Y_t) \sim \text{Ber}(p + q - pq)$ $\Rightarrow \text{prob either or both have arrival at time } t$ <p>Splitting</p> <p>flip a coin with prob q</p> <p>A-Ber(qp)</p> <p>B-Ber($(1-q)p$)</p> <p>\triangleq these streams are not indie</p>	<p>INTER-ARRIVAL TIMES / R. INCIDENCE</p> <p>we arrive at t^*</p> <p>u, v are each $\text{Exp}(\lambda)$ away from t^*</p> <p>$\Rightarrow E[V - U]$ is twice the expectation of $\text{Exp}(\lambda)$</p>	<p>req. iid, $E[X_i] < \infty$ and $\text{Var}(X_i) < \infty$</p> $\sqrt{n} \frac{\overline{X}_n - \mu}{\sigma} \xrightarrow{\text{alt}} N(0, 1)$ $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\text{alt}} N(0, \sigma^2)$ <p>CLT</p>										
<p>COVARIANCE</p> $\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$ <p>Direction</p> $\text{Cov}(X, Y) > 0 \text{ same sign}$ <p>If Indie</p> $\text{Cov}(X, Y) = 0$ <p>\triangleq inverse not usually true but true for Gaussians:</p> $\text{Cov}(X, Y) = 0 \rightarrow X, Y \sim N \text{ indie}$ <p>Properties</p> $\text{Cov}(X, X) = \text{Var}(X)$ $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$ $\text{Cov}(aX + b, Y) = a\text{Cov}(X, Y)$ $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$ <p>\triangleq don't use the above for multiplications!</p>	<p>INTER-ARRIVAL TIMES / R. INCIDENCE</p> <p>TREE</p>	<p>Multiple Engine Example</p> <p>3 engines with death rate λ_e rate until 1st dies is $\lambda = 3\lambda_e$ then rate until 2nd dies $\lambda = 2\lambda_e$</p> <p>Min</p> $P(\min\{X, Y, Z\} \geq t) = P(X \geq t, Y \geq t, Z \geq t) = e^{-3\lambda t}$ <p>\Rightarrow have 3 merged Poissons and want to know first arrival</p> $\Rightarrow \min\{X, Y, Z\} \text{ is Exp}(3\lambda)$ <p>Max</p> $E[\min\{X, Y, Z\}] = \frac{1}{3\lambda}$ <p>LLN</p> <p>req. iid and $E[X_i] < \infty$</p> $\overline{X}_n = \frac{1}{n} \sum^n X_i = E[X]$ $E[\overline{X}_n^2] = \text{Var}(\overline{X}_n) + (E[\overline{X}_n])^2$ <p>\triangleq because \overline{X}_n is a RV like any other</p>	<p>SLUTSKY TH.</p> <p>$T_n \rightarrow T$ and $U_n \rightarrow u$</p> <p>T is r.v. and u is real</p> $T_n + U_n \rightarrow T + u$ $T_n U_n \rightarrow Tu$ $\frac{T_n}{U_n} \rightarrow \frac{T}{u}$ <p>QUANTILES</p> $P(X \leq q_\alpha) = 1 - \alpha$ $\alpha = .1 \rightarrow q_\alpha \text{ is 90th percentile}$ $P(Z > 1.96) = 0.05$ <table border="1" data-bbox="1370 923 1706 1035"> <tr> <td>α</td> <td>2.5%</td> <td>5%</td> <td>7.5%</td> <td>10%</td> </tr> <tr> <td>q_α</td> <td>1.96</td> <td>1.65</td> <td>1.44</td> <td>1.28</td> </tr> </table>	α	2.5%	5%	7.5%	10%	q_α	1.96	1.65	1.44	1.28
α	2.5%	5%	7.5%	10%									
q_α	1.96	1.65	1.44	1.28									
<p>CORRELATION COEF.</p> $\text{Cov}(X, Y) = \frac{E[XY] - E[X]E[Y]}{\sigma_X \sigma_Y}$ <p>measures linear rel</p> <p>FRESH START/MEMORYLESSNESS</p> <p>Exponential</p> $f_{X X>t}(x x > t) = f_X(x)$ <p>Bernouilli/Poisson</p> $P(A B) = P(A)$ <p>i.e. prob of two arrivals (A) after 1 arrival (B) = prob of 2 arrivals (A)</p>	<p>MISC</p> $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \exp(\lambda)$ <p>e limits</p> $\lim_{n \rightarrow \infty} \left(1 - \frac{t}{n}\right)^n = e^{-t}$ $\lim_{n \rightarrow \infty} \left(1 + \frac{t}{n}\right)^n = e^t$ <p>MIN/MAX</p> $P(\max_i X_i > x) = 1 - P(\max_i X_i \leq x) = 1 - [P(X_i \leq x)]^n$ $P(\min_i X_i > x) = [P(X_i > x)]^n = [1 - P(X_i \leq x)]^n$ $P(\min_i X_i < x) = 1 - P(\min_i X_i \geq x)$	<p>LIKELIHOODS</p> <p>Bernouilli $p^{\sum^n X_i} (1-p)^{n - \sum^n X_i}$</p> <p>Poisson $\frac{\lambda^{\sum X_i}}{x_1! \dots x_n!} \exp(-n\lambda)$</p> <p>Gaussian $\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2\right)$</p> <p>Exponential $\lambda^n \exp(-\lambda \sum X_i)$</p> <p>Uniform $\frac{1}{b^n} \mathbf{1}\{ \max_i X_i \leq b \}$</p> <p>$\triangleq a=0 \text{ here}$</p>	<p>IDENTIFIABILITY</p> <p>θ identifiable iff mapping $\theta \in \Theta \rightarrow P_\theta$ is injective (injective: $\theta \neq \theta' \Rightarrow P_\theta \neq P_{\theta'}$)</p> <p>ESTIMATORS</p> <p>Asym. normal if</p> $\sqrt{n}(\widehat{\theta}_n - \theta) \rightarrow N(0, \sigma^2)$ <p>Consistency</p> $\widehat{\theta}_n \rightarrow \theta \text{ as } n \rightarrow \infty$ <p>Bias</p> $\text{bias}(\widehat{\theta}_n) = E[\widehat{\theta}_n] - \theta$ <p>Quadratic Risk</p> $R(\widehat{\theta}_n) = E[(\widehat{\theta}_n - \theta)^2]$ <p>Confidence Interval level $1 - \alpha$</p> <p>conf.int. can't depend on unknown</p> $P\left(\overline{X}_n - q\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n + q\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$ <p>UNBIASED ESTIMATOR</p> <p>we want $\text{bias}[\widehat{\theta}_n] = 0$</p> <p>find $\widehat{\theta}_n$ and use linear property of expectations to create a new estimator such that</p> $E[\widehat{\theta}_n] = cE[\widehat{\theta}_n] = \theta$ <p>1D DELTA METHOD</p> <p>g: cont. differentiable</p> $\sqrt{n}(Z_n - \theta) \rightarrow N(0, \sigma^2)$ $\sqrt{n}(g(Z_n) - g(\theta)) \rightarrow N(0, (g'(\theta))^2 \cdot \sigma^2)$ <p>P-VALUE</p> <p>\triangleq is a level α</p> <p>what is the probability of observing a result more extreme than this one under H_0?</p> <p>\triangleq low p-value is bad $\rightarrow H_0$ is unlikely</p> <p>i.e. $P(\hat{a} \geq \hat{a}_{\text{obs}})$</p> <p>KL DIVERGENCE</p> $KL(P_\theta, P_{\theta'}) = \sum_{x \in E} p_\theta(x) \log \left(\frac{p_\theta(x)}{p_{\theta'}(x)} \right)$ $KL(P_\theta, P_{\theta'}) = \int_E f_\theta(x) \log \left(\frac{f_\theta(x)}{f_{\theta'}(x)} \right) dx$ <p>Properties</p> <p>not symmetric</p> <p>not negative definite</p> <p>triangle ineq</p>										

TESTS

fails to reject H_0 does not mean accepting H_0

Errors

test reality	H_0	H_1
H_0	✓	type 1 error (reject when shouldn't)
H_1	type 2 error (fail to reject when should)	✓

level α

max type 1 error rate

higher $\alpha \rightarrow$ more likely to reject H_0

power β

$$\pi_\psi = \inf_{\theta \in \Theta} (1 - \beta_\psi(\theta))$$

example 2 sided

coin $H_0: p = \frac{1}{2}$ and $H_1: p \neq \frac{1}{2}$

$$\psi = 1 \left\{ \sqrt{n} \left| \bar{X}_n - \frac{1}{2} \right| > \frac{q_\alpha}{2} \right\}$$

stats diff between X and Y? (ex.)

$\bar{X}_n \sim N(\mu_1, \sigma_1^2)$ and $\bar{Y}_n \sim N(\mu_2, \sigma_2^2)$

$H_0: \mu_1 = \mu_2$ and $H_1: \mu_1 \neq \mu_2$

$$\sqrt{n} \frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\sigma_1^2 + \sigma_2^2}} \sim N(0, 1)$$

single-sided

evaluate H_0 at boundary (see part c here)

$H_0 \mu \geq \sigma$ and $H_1 \mu < \sigma$

boundary is $\mu = \sigma$ for $g(\theta)$ or θ

TOTAL VARIATION DISTANCE

max dist between two distros

ΔE is joint set of values of RVs

$$TV(P_\theta, P_{\theta'}) = \frac{1}{2} \sum_{x \in E} |p_\theta(x) - p_{\theta'}(x)|$$

$$TV(P_\theta, P_{\theta'}) = \frac{1}{2} \int_{-\infty}^{\infty} |f_\theta(x) - f_{\theta'}(x)| dx$$

Properties

symmetric: $TV(P_\theta, P_{\theta'}) = TV(P_{\theta'}, P_\theta)$

positive: $0 \leq TV \leq 1$

definite: if $TV(P_\theta, P_{\theta'}) = 0$ then $P_\theta = P_{\theta'}$

triangle ineq:

$$TV(P_\theta, P_{\theta'}) \leq TV(P_\theta, P_{\theta''}) + TV(P_{\theta''}, P_{\theta'})$$

if disjoint: $TV = 1$

if same: $TV = 0$

MAXIMIZATION

global extremes on range

test critical points and end points

min/max

$h''(x) \leq 0 \rightarrow$ concave, maximum, h' decr.

$h''(x) < 0 \rightarrow$ concave, global max, h' decr.

$h''(x) \geq 0 \rightarrow$ convex, minimum, h' incr.

MV min/max

$X^T H h(\theta) X \leq 0$ concave, max

+1 top diag: convex, minimum

MLE

minimizes KL divergence

$$\hat{\theta}_n^{MLE} = \arg \max_{\theta \in \Theta} \log(L)$$

MLE can be Biased

function must be cont. diff. to use derivative to find extremums. use a plot and think if not

Consistency and Asym. Norm.

if

- param is identifiable
- support of P_θ does not depend on θ
- θ^* is not at boundary
- $I(\theta)$ is invertible
- more stuff

then

consistent: $\hat{\theta}_n^{MLE} \rightarrow \theta^*$

A. normal: $\sqrt{n}(\hat{\theta}_n^{MLE} - \theta^*) \rightarrow N(0, I(\theta^*)^{-1})$

Process to find extremum

- get l_n
- find crits with $l_n'(\theta) = 0$
- check if crits are local min/max
- check values at endpoints

METHOD OF MOMENTS

$$\hat{m}_k = \bar{X}_n^k = \frac{1}{n} \sum X_i^k$$

$$\text{LLN } \hat{m}_k \rightarrow m_k(\theta) = E_\theta[X_1^k]$$

$$\text{ASYM NORM } \sqrt{n}(\hat{\theta} - \theta) \rightarrow N(0, \Gamma(\theta))$$

$$\Gamma(\theta) = \left[\frac{\delta M^{-1}}{\delta \theta} \right]^T \Sigma(\theta) \left[\frac{\delta M^{-1}}{\delta \theta} \right]$$

finding $\hat{\theta}$

write θ as function $E[X]$, $E[X^2] \dots$

then sub for \bar{X}_n , \bar{X}_n^2

M-ESTIMATION

Lecture 12, tab 2

FISHER INFORMATION

use ONE observation

not well defined if support depends on unknown (shifted exp)

$\Delta l'(\theta)$ must exist

$$I(\theta) = \text{Var}(l'(\theta)) = -E[l''(\theta)]$$

Δ the $E[\cdot]$ is of the observation X and not the unknown! $E[\theta X] = \theta E[X]$

χ^2 DISTRO

distro of sum of $Z_i \sim N(0, 1)$

$$E[V] = d$$

$$\text{Var}(V) = 2d$$

COCHRAN'S TH.

$$\frac{nS_n}{\sigma^2} \sim \chi^2_{n-1} \text{ or } nS_n \sim \frac{\sigma^2}{n} \chi^2_{n-1}$$

t DISTRO

for small nb of Gaussian samples w/

$$Z \sim N(0, 1) \text{ and } V \sim \chi^2_d \text{ and } \text{SampleVar} = \frac{V}{d}$$

$$\frac{Z}{\sqrt{\frac{V}{d}}} \Delta Z \text{ and } V \text{ must be indie}$$

t TEST

requires Gaussian samples

is pivotal (q in tables)

test is non-asymptotic

one sample two-sided

$H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$

$$T_n = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sqrt{\hat{S}_n}} = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sqrt{\frac{\sum (X_i - \bar{X}_n)^2}{n-1}}} \sim t_{n-1}$$

$$\psi_\alpha = \mathbf{1}\{T_n > q_\alpha\}$$

one sample one-sided

$H_0: \mu \leq \mu_0$ vs $H_1: \mu > \mu_0$

$$T_n = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sqrt{\hat{S}_n}} \sim t_{n-1}$$

$$\psi_\alpha = \mathbf{1}\{T_n > q_\alpha\}$$

two sample

$$\bar{X}_n \sim N\left(\Delta_d, \frac{\sigma_d^2}{n}\right) \text{ and } \bar{Y}_n \sim N\left(\Delta_c, \frac{\sigma_c^2}{m}\right)$$

$$\bar{X}_n - \bar{Y}_m \sim N\left(\Delta_d - \Delta_c, \frac{\sigma_d^2}{n} + \frac{\sigma_c^2}{m}\right)$$

$$\text{where } N \text{ according to Welch-Satter:}$$

$$N = \frac{\left(\frac{\sigma_d^2}{n} + \frac{\sigma_c^2}{m}\right)^2}{\frac{\sigma_d^4}{n^2} + \frac{\sigma_c^4}{m^2}} \geq \min(n, m)$$

WALD'S TEST

- test is asymptotic
- not invariant to change in rep of H_0

only req est of unrestricted model, lower computation

Δ MLE conditions must be satisfied

$H_0: \theta = \theta_0$ vs $H_1: \theta \neq \theta_0$ for $\theta \in \mathbb{R}^d$

$$n(\hat{\theta}^{MLE} - \theta_0)^T I(\hat{\theta}^{MLE})(\hat{\theta}^{MLE} - \theta_0) \rightarrow \chi^2_d$$

equivalently

$$T_n = \left\| \sqrt{n}(\hat{\theta}_0 - \theta_0) \right\|^2 \rightarrow \chi^2_d$$

which gives test $\psi_\alpha = \mathbf{1}\{T_n > q_\alpha\}$

where q_α is the $(1 - \alpha)$ -quantile of χ^2_d

X_i iid $\Theta \in \mathbb{R}^d$

$$H_0: (\theta_{r+1}, \dots, \theta_d) = (\theta_{r+1}^{(0)}, \dots, \theta_d^{(0)})$$

$$T_n = 2 \left(l_n(\hat{\theta}_n^{MLE}) - l_n(\theta_n^{(0)}) \right) = -2 \ln \left(\frac{L_{H_0}}{L_{H_1}} \right)$$

Δ same n for both likelihoods

wilk Th. (dim. ex. @ 16:00)

assuming H_0 is true and MLE conditions. is asymptotic

$$T_n \rightarrow \chi^2_{d-r}$$

$$\psi_\alpha = \mathbf{1}\{T_n > q_\alpha\}$$

CATEGORICAL LIKELIHOOD

i.e. are Zodiac signs uniformly distributed?

$$p_0 = \left(\frac{1}{12}, \frac{1}{12}, \dots \right)$$

$L_n = p_1^N - 1 \dots p_k^N - 1$

$$N_j = \#\{X_i = a_j\}$$

$$\hat{p} \rightarrow \hat{p}_j = \frac{N_j}{n} \text{ prob of obs. outcome j}$$

$$p_j = P(X = a_j) = \prod_i \mathbf{1}_{\{a_i = a_j\}}$$

χ^2 TEST

$H_0: \vec{p} = \vec{p}^0$ vs $H_1: \vec{p} \neq \vec{p}^0$

$$T_n = n \sum_{j=0}^k \left[\frac{(\hat{p}_j - p_j^0)^2}{p_j^0} \right] \rightarrow \chi^2_{k-1}$$

where k is nb of categories

χ^2 TEST FOR FAMILY OF DIST

$H_0: p \in \{\text{Bin}(k, \theta)\}_{\theta \in \Theta}$ vs $H_1: p \notin \{\text{Bin}(k, \theta)\}_{\theta \in \Theta}$

$$T_n = n \sum_{j=0}^k \frac{\left(\frac{N_j}{n} - f_{\hat{\theta}}(j) \right)^2}{f_{\hat{\theta}}(j)} \rightarrow \chi^2_{k(d-1)-d-1}$$

$\Delta k-d-1$ if we start at j=1

$\theta \in \mathbb{R}^d$

f_θ is PMF of $\text{Bin}(k, \theta)$

$\hat{\theta}$ is MLE here

EMPIRICAL CDF

$$F_n(t) = \frac{1}{n} \sum \mathbf{1}\{X_i \leq t\}$$

it is discontinuous

$$\sqrt{n}(F_n(t) - F(t)) \rightarrow N(0, F(t)(1 - F(t)))$$

DONSKER'S TH.

if F cont:

$$\sqrt{n} \max_{t \in \mathbb{R}} |F_n(t) - F(t)| \rightarrow \max_{0 \leq t \leq 1} |B(t)|$$

where B is Brownian bridge

KS TEST (example)

X_i : real RV with unk CDF

$H_0: F = F^0$ vs $H_1: F \neq F^0$

$$\delta_\alpha^{KS} = \mathbf{1}\{T_n > q_\alpha\}$$

$$= \mathbf{1}\left\{ \max_{t \in \mathbb{R}} \sqrt{n}|F_n(t) - F(t)| > q_\alpha \right\}$$

p-value $P(Z > T_n \mid T_n)$

computation

$$\frac{T_n}{\sqrt{n}} = \max_{1 \leq i \leq n} \left[\max \left(\left| F^0(x_i) - \frac{i}{n} \right|, \left| F^0(x_i) - \frac{i-1}{n} \right| \right) \right]$$

needs tables (pivotal statistic)

KL TEST (example)

is my data Gaussian?

more likely to reject than KS test

$$\max_{t \in \mathbb{R}} |F_n(t) - \Phi_{\hat{\mu}, \hat{\sigma}^2}(t)|$$

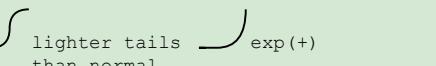
QQ PLOT (example 1, 2)

$$F_n^{-1}\left(\frac{i}{n}\right) = X_i \quad (F_n \text{ is sample CDF, } F \text{ is th.})$$

points are $\left(F_n^{-1}\left(\frac{1}{n}\right), x_1\right), \left(F_n^{-1}\left(\frac{2}{n}\right), x_2\right) \dots$

to find inverse F^{-1} : "what input value to F gives output value t. we are looking

for input value to F that gives $\frac{1}{n}$ "



MARKOV INEQUALITY

$X \geq 0$ and $a > 0$

$$P(X \geq a) \leq \frac{E[X]}{a}$$

CHEBYSHEV INEQUALITY (link)

probability of estimate of mean deviating from true mean by more than C

$$P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

CONVERGENCE IN PROBABILITY

a seq. converges to a in probability if:

$$\lim_{n \rightarrow \infty} P(|X_n - \mu| \geq \epsilon) = 0$$

another way to show convergence in prob

is to determine expectation and variance. if $\text{Var} \rightarrow 0$ then convergence

properties

if g is continuous then $g(X_n) \rightarrow g(a)$

$X_n + Y_n \rightarrow a + b$

but $E[X_n]$ doesn't need to converge to a

ESTIMATE BINOMIAL WITH NORMAL

PMF of # success in n trials w/p p approximates $N(np, np(1-p))$

with

$$'P(X = 19)' = P(18.5 \leq X \leq 19.5)$$

MOIVRE LAPLACE CORRECTION

when estimating an integer R.V. with the CLT, can do the "1/2 correction":

$$P(S_n \leq 21) \rightarrow P(S_n \leq 21.5)$$

is estimator consistent?

check lim as $n \rightarrow \infty$ against estimator

is estimator asym. normal?

start with CLT definition, then put in the estimator. also get aVar like this.

see examples.

<p>BAYESIAN STATS</p> $\pi(\theta X_1 \dots X_n) = \frac{\pi(\theta) L_n(X_1 \dots X_n \theta)}{\int_{\Theta} \pi(\theta) L_n(X_1 \dots X_n \theta)}$ $\propto \pi(\theta) L_n(X_1 \dots X_n \theta)$ <p>conjugate prior if post. distro. same as prior distro.</p> <p>improper prior i.e. uniform $\pi(\theta) = 1$, not a valid distro</p> <p>Jeffrey's prior non-informative prior, not always improper. reflects no prior belief, only stats model</p> $\pi_J(\theta) \propto \sqrt{\det I(\theta)}$ <p>reparam. invariance we have Jeff prior for θ, want $\eta = \Phi(\theta)$</p> <ul style="list-style-type: none"> · replace θ with $\Phi^{-1}(\eta)$ · multiply by $\frac{d\theta}{d\eta} = \frac{1}{\Phi'(\theta)}$ <p>confidence region $P(\theta \in \mathbb{R} X_1 \dots X_n) = 1 - \alpha$</p>	<p>BAYES ESTIMATOR</p> <p>mean of posterior also known as LMS "conditional expectation" $E[\Theta X = x]$</p> <p>⚠ MUST USE ACTUAL POSTERIOR, not the prop. one if we calculate it like below, else we may also use mean of the distribution if i.e. Beta without having to calculate denominator</p> $\hat{\theta}^* = \int_{\Theta} \theta \pi(\theta X_1 \dots X_n) d\theta$ <p>aVar = $I^{-1}(\theta)$ of distro sampled</p> <p>properties of LMS estimation error</p> <p>let $\tilde{\Theta} = E[\Theta X]$ and error $\tilde{\Theta} = \hat{\Theta} - \theta^*$</p> <ul style="list-style-type: none"> · $E[\tilde{\Theta} X = x] = 0$ · $cov(\tilde{\Theta}, \hat{\Theta}) = 0$ · $Var(\Theta) = Var(\hat{\Theta}) + Var(\tilde{\Theta})$ <p>conditional MSE of LMS estimator</p> $E[(\Theta - \hat{\Theta})^2 X = x] = Var(\Theta X = x)$	<p>MV LINEAR REGRESSION (STATS)</p> $\vec{Y} = \mathbb{X}\beta^* + \vec{\varepsilon}$ $\vec{\beta} \in \mathbb{R}^p, \vec{Y} \in \mathbb{R}^n, \mathbb{X} \in \mathbb{R}^{n \times p}$ <p>LSE (same as Bayes estimator)</p> $\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^p} \ \vec{Y} - \mathbb{X}\beta\ ^2$ $\hat{\beta} = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \vec{Y}$ <p>$Rank(\mathbb{X}) = p$ and need $n \geq p$ for this to work</p> <p>assumptions</p> <ul style="list-style-type: none"> · \mathbb{X} is deterministic, rank=p · ε_i are iid · $\varepsilon \sim N(0, \sigma^2 I_n)$ $\Rightarrow Y \sim N_n(\mathbb{X}\beta^*, \sigma^2 I_n)$ $\Rightarrow I(\beta) = \frac{1}{\sigma^2} \mathbb{X}^T \mathbb{X}$ <p>properties of LSE</p> <ul style="list-style-type: none"> · LSE is MLE in homoscedastic Gaussian case · $\hat{\beta} \sim N_p(\beta^*, \sigma^2 (\mathbb{X}^T \mathbb{X})^{-1})$ ⚡ asym · quadratic risk: $E[\ \hat{\beta} - \beta\ ^2] = \sigma^2 \text{trace}((\mathbb{X}^T \mathbb{X})^{-1})$ · prediction error: $E[\ Y - \mathbb{X}\hat{\beta}\ ^2] = \sigma^2(n - p)$ · unbiased estimator: $\sigma^2 = \frac{\ Y - \mathbb{X}\hat{\beta}\ ^2}{n - p} = \frac{1}{n - p} \sum \varepsilon^2$ <p>theorems</p> <ul style="list-style-type: none"> · $(n - p) \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}$ · \hat{B} and $\hat{\sigma}^2$ are orthogonal and indie 	<p>WOLFRAM</p> <p>Probability $x > 4.03$, Chi Squared Distribution degrees of freedom 1 $CDF[NormalDistribution[2, 1], 0.65]$ ⚡ CDF uses STANDARD DEVIATION $\text{Quantile}[\text{ChiSquareDistribution}[1], 0.95]$ $\text{Round}[5.15517, 0.001]$</p> <p>1 PARAM CANON EXP FAMILY (ex)</p> $f_\theta(y) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y, \theta)\right)$ <p>θ is canon. param ϕ (dispersion), b and c known $b(\theta)$ is log partition $E[Y] = b'(\theta)$ $Var(Y) = b''(\theta)\phi$</p> <p>⚠ we can do a substitution i.e. $\theta = \frac{1}{\lambda}$ to express. then substitute again to express as λ</p> <p>linear transformations of these are also canon.</p> <p>canon link links $\mu(x)$ to canon param θ: $g(\mu(x)) = \theta = (b')^{-1}(\mu(x))$ if $\phi > 0$ canon link is strictly increasing</p> <p>GLM MODEL</p> <p>$\vec{Y} = (Y_1, \dots, Y_n)$ and $\mathbb{X} = (X_1, \dots, X_n)$ $\mu_i = E[Y_i X_i]$ is related to canonical param θ_i via $\mu_i = b'(\theta_i)$ μ_i depends linearly on the covariates through link function g: $g(\mu_i) = X_i^T \beta$</p> <p>using predictor use mean function in table below once we have $\hat{\beta}$</p> <p>asymptotic normality $\hat{\beta}$ is asym normal finding β MLE/Gradient Descent</p>																																																
<p>BAYESIAN STATS - NORMALS</p> $f_X(x) = c \exp(-(\alpha x^2 + \beta x + \gamma))$ $\mu = -\frac{\beta}{2\alpha} \text{ and } \sigma^2 = \frac{1}{2\alpha}$ <p>the peak is min. of exponent: · derive exponent and set to 0</p> <p>$\hat{\Theta}_{MAP} = \hat{\Theta}_{LMS} = E[\Theta X = x]$ (in general this is true if posterior is unimodal and symmetric)</p> <p>MAP</p> $\hat{\theta}_{MAP} = \arg \max_{\theta} \pi(\theta X_1 \dots X_n)$ $= \arg \max_{\theta} \pi(\theta) L_n(X_1 \dots X_n \theta)$ <p>⚠ look at posterior PDF/PMF and ask "which <u>actual</u> possible values of θ make this result most likely, i.e. the mode i.e. is $\theta_1 - \hat{\theta}_{Bayes} > \theta_2 - \hat{\theta}_{Bayes}$</p> <p>⚠ if discrete, MAP is in set of possible values find MAP continuous take derivative, find critical points, maximum</p>	<p>LLMS / LINEAR REGRESSION</p> <p>unknown Θ, observation X</p> $\hat{\Theta} = aX + b$ <p>minimises $E[(\Theta - aX - b)^2]$</p> $a = \frac{\text{Cov}(\Theta, X)}{\text{Var}(X)}$ $b = E[\Theta] - aE[X]$ <p>⚠ if all vars normals then LMS=LLMS</p> <p>MSE</p> $MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = Var(\hat{\theta}) + (Bias(\hat{\theta}))^2$ <p>Gaussian</p> $MSE(\bar{X}_n) = E[(\bar{X}_n - \mu)^2] = \left(\frac{\sigma}{\sqrt{n}}\right)^2$ $MSE(\bar{S}_n) = \frac{2}{n-1} \sigma^4$ $MSE(S_n) = \frac{2n-1}{n^2} \sigma^4$ <p>LINEAR REGRESSION FUNCTION</p> $E[Y X = x] = \mu(x) = \int y h(y x) dy = X^T \beta$	<p>BONFERRONI'S TEST (ex.)</p> <p>test whether group of explanatory vars is significant FWER $\leq \alpha$</p> <p>⚠ non asymptotic test</p> <p>$H_0: \beta_j = 0 \ \forall j \in S$ where $S \subseteq \{1, \dots, p\}$</p> <p>$H_1: \exists j \in S$ where $\beta_j \neq 0$</p> <p>$R_{S,\alpha} = \bigcup_{j \in S} R_{j,\frac{\alpha}{k}}$ (OR statement!)</p> <p>where k is # in S, and $\frac{\alpha}{k}$ usually passed to a 2 sided test so that final quantile may be $q_{\frac{\alpha}{2k}}$</p> $\psi = 1 \left\{ \frac{\max(\hat{\beta}_1 , \hat{\beta}_2 , \dots)}{\sqrt{Var(\hat{\beta}_j)}} > q_{\frac{\alpha}{2k}} \right\}$	<p>CANON PARAMETER</p> $\theta = a + bX = \mathbb{X}\beta = g(\mu)$ <p>here μ is the param of our distro, and θ is the canon param</p> <p>Probability distribution</p> $\mu = g^{-1}(\theta)$ <p>MULTIPLE HYPOTHESIS TESTING (see @ 9:24)</p> <p>family-wise error rate</p> <p>FWER = P(at least one false significant result) $\leq \alpha$ use Bonferroni's test $= 1 - P(V = 0) = 1 - 0.95^{100} \approx 0.99$</p> <p>very restrictive reject when $m \cdot p - \text{value} \leq \alpha$</p> <p>false discovery rate</p> <p>FDR = expected fraction of false significant results among all significant results $\leq \alpha$</p> <p>Holm-Bonferroni correction Bonferroni-Hochberg correction</p>																																																
<p>LINEAR REGRESSION (STATS)</p> <p>this describes the practical model. LLMS in Prob describes theory.</p> <p>⚠ nb: stats and prob flip the a, b like theoretical model but assume some Gaussian noise</p> $Y_i = a^* + b^* X_i + \varepsilon_i$ <p>use least squares to find estimators</p> $\min \sum (Y_i - a - bX_i)^2$ $\hat{a} = \bar{Y} - b\bar{X}$ $\hat{b} = \frac{\bar{XY} - \bar{X}\bar{Y}}{\bar{X}^2 - \bar{X}^2}$	<p>SIGNIFICANCE TESTS</p> <p>is j^{th} explanatory variable significant</p> <p>$H_0: \beta_j = 0$ $H_1: \beta_j \neq 0$ (ex. for $\beta_1 = \beta_2$)</p> <p>assume γ_j is j^{th} diagonal coefficient of $(\mathbb{X}^T \mathbb{X})^{-1}$ ($\gamma_j > 0$)</p> $\Rightarrow T_n = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 - \gamma_j}} \sim t_{n-p}$ $\Rightarrow R_{j,\alpha} = \left\{ \left T_n^{(j)} \right > q_{\frac{\alpha}{2}}(t_{n-p}) \right\}$	<table border="1"> <thead> <tr> <th>Distribution</th> <th>Support of distribution</th> <th>Typical uses</th> <th>Link name</th> <th>Canon Link function, $\mathbb{X}\beta = g(\mu)$</th> <th>Mean function</th> </tr> </thead> <tbody> <tr> <td>Normal</td> <td>real: $(-\infty, +\infty)$</td> <td>Linear-response data</td> <td>Identity</td> <td>$\mathbb{X}\beta = \mu$</td> <td>$\mu = \mathbb{X}\beta$</td> </tr> <tr> <td>Exponential</td> <td>real: $(0, +\infty)$</td> <td>Exponential-response data, scale parameters</td> <td>Negative inverse</td> <td>$\mathbb{X}\beta = -\mu^{-1}$</td> <td>$\mu = -(\mathbb{X}\beta)^{-1}$</td> </tr> <tr> <td>Gamma</td> <td>real: $(0, +\infty)$</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Inverse Gaussian</td> <td>real: $(0, +\infty)$</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Poisson</td> <td>integer: $0, 1, 2, \dots$</td> <td>count of occurrences in fixed amount of time/space</td> <td>Log</td> <td>$\mathbb{X}\beta = \ln(\mu)$</td> <td>$\mu = \exp(\mathbb{X}\beta)$</td> </tr> <tr> <td>Bernoulli</td> <td>integer: $\{0, 1\}$</td> <td>outcome of single yes/no occurrence</td> <td></td> <td>$\mathbb{X}\beta = \ln\left(\frac{\mu}{1-\mu}\right)$</td> <td>$\mu = \frac{\exp(\mathbb{X}\beta)}{1+\exp(\mathbb{X}\beta)}$</td> </tr> <tr> <td>Binomial</td> <td>integer: $0, 1, \dots, N$</td> <td>count of # of "yes" occurrences out of N yes/no occurrences</td> <td>Logit</td> <td>$\mathbb{X}\beta = \ln\left(\frac{\mu}{n-\mu}\right)$</td> <td>$\mu = \frac{\exp(\mathbb{X}\beta)}{1+\exp(\mathbb{X}\beta)}$</td> </tr> </tbody> </table>	Distribution	Support of distribution	Typical uses	Link name	Canon Link function, $\mathbb{X}\beta = g(\mu)$	Mean function	Normal	real: $(-\infty, +\infty)$	Linear-response data	Identity	$\mathbb{X}\beta = \mu$	$\mu = \mathbb{X}\beta$	Exponential	real: $(0, +\infty)$	Exponential-response data, scale parameters	Negative inverse	$\mathbb{X}\beta = -\mu^{-1}$	$\mu = -(\mathbb{X}\beta)^{-1}$	Gamma	real: $(0, +\infty)$					Inverse Gaussian	real: $(0, +\infty)$					Poisson	integer: $0, 1, 2, \dots$	count of occurrences in fixed amount of time/space	Log	$\mathbb{X}\beta = \ln(\mu)$	$\mu = \exp(\mathbb{X}\beta)$	Bernoulli	integer: $\{0, 1\}$	outcome of single yes/no occurrence		$\mathbb{X}\beta = \ln\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{\exp(\mathbb{X}\beta)}{1+\exp(\mathbb{X}\beta)}$	Binomial	integer: $0, 1, \dots, N$	count of # of "yes" occurrences out of N yes/no occurrences	Logit	$\mathbb{X}\beta = \ln\left(\frac{\mu}{n-\mu}\right)$	$\mu = \frac{\exp(\mathbb{X}\beta)}{1+\exp(\mathbb{X}\beta)}$	
Distribution	Support of distribution	Typical uses	Link name	Canon Link function, $\mathbb{X}\beta = g(\mu)$	Mean function																																														
Normal	real: $(-\infty, +\infty)$	Linear-response data	Identity	$\mathbb{X}\beta = \mu$	$\mu = \mathbb{X}\beta$																																														
Exponential	real: $(0, +\infty)$	Exponential-response data, scale parameters	Negative inverse	$\mathbb{X}\beta = -\mu^{-1}$	$\mu = -(\mathbb{X}\beta)^{-1}$																																														
Gamma	real: $(0, +\infty)$																																																		
Inverse Gaussian	real: $(0, +\infty)$																																																		
Poisson	integer: $0, 1, 2, \dots$	count of occurrences in fixed amount of time/space	Log	$\mathbb{X}\beta = \ln(\mu)$	$\mu = \exp(\mathbb{X}\beta)$																																														
Bernoulli	integer: $\{0, 1\}$	outcome of single yes/no occurrence		$\mathbb{X}\beta = \ln\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{\exp(\mathbb{X}\beta)}{1+\exp(\mathbb{X}\beta)}$																																														
Binomial	integer: $0, 1, \dots, N$	count of # of "yes" occurrences out of N yes/no occurrences	Logit	$\mathbb{X}\beta = \ln\left(\frac{\mu}{n-\mu}\right)$	$\mu = \frac{\exp(\mathbb{X}\beta)}{1+\exp(\mathbb{X}\beta)}$																																														