## Distributions

**Normal:** $X \sim N(\mu, \sigma^2)$, pdf: $f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, mean: $\mu$, variance: $\sigma^2$. <u>SM</u>: Box-Muller Transform: If $U_1, U_2 \sim Uniform(0,1)$ i.i.d., then $Z_1 = \sqrt{-2\ln U_1}\cos(2\pi U_2)$ and $Z_2 = \sqrt{-2\ln U_1}\sin(2\pi U_2)$ are i.i.d. $N(0,1)$. To get $N(\mu, \sigma^2)$, use $X = \sigma Z + \mu$

**Bernoulli:** $X \sim Bern(p)$, pmf: $P(X = 1) = p$, $P(X = 0) = 1 - p$, mean: $p$, variance: $p(1-p)$ <u>SM</u>: If $U \sim Uniform(0,1)$, then $X = 1$ if $U \leq p$, else $X = 0$

**Binomial:** $X \sim Bin(n,p)$, pmf: $P(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$, mean: $np$, variance: $np(1-p)$: Note: $Bin(n,p) = \sum_{i=1}^{n} Bern(p)$ <u>SM</u>: If $U_i \sim Uniform(0,1)$ i.i.d., then $X = \sum_{i=1}^{n} I(U_i \leq p)$

**Multinomial** $\sim Mult(n, p_1, p_2, ..., p_k)$, pmf: $P(X_1 = x_1, X_2 = x_2, ..., X_k = x_k) = \frac{n!}{x_1!x_2!...x_k!}p_1^{x_1}p_2^{x_2}...p_k^{x_k}$, mean: $E[X_i] = np_i$, variance: $Var(X_i) = np_i(1-p_i)$, covariance: $Cov(X_i, X_j) = -np_ip_j$ for $i \neq j$ <u>SM</u>: If $U_i \sim Uniform(0,1)$ i.i.d., then for each $U_i$, assign it to category $j$ if $\sum_{m=1}^{j-1} p_m < U_i \leq \sum_{m=1}^{j} p_m$, then $X_j$ is the count of assignments to category $j$

**Exponential**: $X \sim Exp(\lambda)$, pdf: $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$, mean: $\frac{1}{\lambda}$, variance: $\frac{1}{\lambda^2}$ <u>SM</u>: If $U \sim Uniform(0,1)$, then $X = -\frac{1}{\lambda}\ln(U)$

**Poisson**: $X \sim Poisson(\lambda)$, pmf: $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ for $k = 0, 1, 2, ...$, mean: $\lambda$, variance: $\lambda$ Note: $Poisson(\lambda)$ with Exponential inter-arrival times <u>SM</u>: $U \sim U(0,1)$ set $f = e^{-\lambda}$, $k = 0$, $F = f$, while $F < u$ do $k = k + 1$, $f = f\frac{\lambda}{k}$, $F = F + f$, return $k$

**Chi-Squared**: $X \sim \chi_k^2$, pdf : $f(x) = \frac{1}{2^{k/2}\Gamma(k/2)}x^{(k/2)-1}e^{-x/2}$ for $x \geq 0$, mean: $k$, variance: $2k$ Note: If $Z_i \sim N(0,1)$ i.i.d., then $\sum_{i=1}^{k} Z_i^2 \sim \chi_k^2$ <u>SM</u> : $z_k \sim N(0,1)$ $x = (z_1 + \lambda)^2 + \sum_{i=2}^{k} z_i^2$

**t-Distribution**: $X \sim t_k$, pdf: $f(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\Gamma(\frac{k}{2})}(1 + \frac{x^2}{k})^{-\frac{k+1}{2}}$, mean: 0 for $k > 1$, variance: $\frac{k}{k-2}$ for $k > 2$ Note: If $Z \sim N(0,1)$ and $V \sim \chi_k^2$ independent, then $\frac{Z}{\sqrt{V/k}} \sim t_k$ <u>SM</u> Use the property of $T \sim N/\sqrt{(X/k)}$ where $N \sim N(0,1)$ and $X \sim \chi_k^2$ independent. Generate $N$ using Box-Muller and $X$ using Chi-Squared SM.

**F-Distribution**: $X \sim F_{d_1,d_2}$, pdf: $f(x) = \frac{\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1+d_2}}}}{xB(\frac{d_1}{2}, \frac{d_2}{2})}$ for $x \geq 0$, mean: $\frac{d_2}{d_2-2}$ for $d_2 > 2$, variance: $\frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$ for $d_2 > 4$ Note: If $U_1 \sim \chi_{d_1}^2$ and $U_2 \sim \chi_{d_2}^2$ independent, then $\frac{(U_1/d_1)}{(U_2/d_2)} \sim F_{d_1,d_2}$ <u>SM</u>: Generate $U_1$ and $U_2$ using Chi-Squared SM and use the property above.

**Gamma**: $X \sim Gamma(\alpha, \beta)$, pdf: $f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}$ for $x \geq 0$, mean: $\alpha\beta$, variance: $\alpha\beta^2$ <u>SM</u>: If $\alpha$ is an integer, generate $\alpha$ i.i.d. $Exp(1/\beta)$ and sum them. If not integer, use acceptance-rejection or other methods.

**Beta**: $X \sim Beta(\alpha, \beta)$, pdf: $f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$ for $0 \leq x \leq 1$, mean: $\frac{\alpha}{\alpha+\beta}$, variance: $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ <u>SM</u>: If $U_1 \sim Gamma(\alpha, 1)$ and $U_2 \sim Gamma(\beta, 1)$ independent, then $X = \frac{U_1}{U_1+U_2} \sim Beta(\alpha, \beta)$

**Inverse CDF**: For any $X \sim F(x)$, if $U \sim Uniform(0,1)$, then $X = F^{-1}(U)$ has distribution $F(x)$. Use when $F^{-1}$ is available in closed form.

## Sampling Methods

**Rejection Sampling**: $X \sim f(x)$, $g(x)$ is proposal distribution with $f(x) \leq Mg(x)$ for all $x$, generate $Y \sim g(y)$ and $U \sim Uniform(0,1)$, accept $Y$ if $U \leq \frac{f(Y)}{Mg(Y)}$, else reject and repeat.

**Importance Sampling**: $X \sim f(x)$ where $f(x) = \frac{h(x)}{\int h(x)dx}$ , $g(x)$ is proposal distribution, generate $X_i \sim g(x)$ i.i.d. for $i = 1, 2, ..., n$, Then sample $X$ from $Y_i$ with weight $w_i = \frac{f(X_i)}{g(X_i)} = \frac{h(x)}{g(x)}$ normalized so that $\sum_{i=1}^{n} w_i = 1$. We can use this for estimating expectations: $E_f[t(X)] = \int t(x)f(x)dx = \int t(x)\frac{f(x)}{g(x)}g(x)dx = E_Z[t(Z)\frac{f(Z)}{g(Z)}]Z \sim g(x) \approx \frac{1}{n}\sum_{i=1}^{n} t(z_i)\frac{f(z_i)}{g(z_i)}$

**Gibbs Sampling**: To sample from joint distribution $f(x, y)$, initialize $X^{(0)}$ and $Y^{(0)}$, then for $i = 1, 2, ..., n$, sample $X^{(i)} \sim f(x|Y^{(i-1)})$ and $Y^{(i)} \sim f(y|X^{(i)})$. The samples $(X^{(i)}, Y^{(i)})$ converge to the joint distribution $f(x, y)$ as $n \to \infty$ Need to do burn-in and thinning to reduce autocorrelation. For multi-dimensions, sample each variable in turn conditioned on the others with full conditional distributions.

## MCMC & Metropolis-Hastings

**MC Method/Integration** : To estimate $I = \int_a^b f(x)dx$, generate $U_i \sim Uniform(a,b)$ i.i.d. for $i = 1, 2, ..., n$, then $\hat{I} = \frac{b-a}{n}\sum_{i=1}^{n} f(U_i)$

**MCMC**: To sample from target distribution $f(x)$, construct a Markov chain with transition kernel $P(x, y)$ such that $f(x)$ is the stationary distribution. Run the chain for a long time and use the samples to estimate expectations. Comnverges requires $\pi$-irreducibility, aperiodic, and invariance distribution.

**Metropolis-Hastings**: To sample from target distribution $f(x)$, choose $g(\cdot|x)$. Init $x_0$. For $i = 1, 2, ..., n$, generate $Y \sim g(\cdot|x_{i-1})$, compute acceptance ratio $\alpha = \min(1, \frac{f(Y)g(x_{i-1}|Y)}{f(x_{i-1})g(Y|x_{i-1})})$, $x_i = Y$ with probability $\alpha$, else $x_i = x_{i-1}$. The samples $x_i$ converge to $f(x)$ as $n \to \infty$.

**Metropolis Algorithm**: Special case of Metropolis-Hastings where $g(y|x) = g(x|y)$ (symmetric proposal). Acceptance ratio simplifies to $\alpha = \min(1, \frac{f(Y)}{f(x_{i-1})})$.

**Independent MH**: Special case of Metropolis-Hastings where $g(y|x) = g(y)$ (independent proposal). Acceptance ratio is $\alpha = \min(1, \frac{f(Y)g(x_{i-1})}{f(x_{i-1})g(Y)})$.

<div align="center">Correlation Coefficient</div>

**Pearson Correlation Coefficient**: For random variables $X$ and $Y$, $\rho_p(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{E[(X-E[X])(Y-E[Y])]}{\sqrt{E[(X-E[X])^2]E[(Y-E[Y])^2]}}$ Used for measuring linear relationship between variables. Sample: $\hat{\rho}_p = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}}$

**Spearman's Rank Correlation Coefficient**: For random variables $X$ and $Y$, $\rho_s(X,Y) = \rho_p(F_X(z), F_Y(z)) = 12 \int \int F_X F_Y f(x,y)dxdy - 3$ where $F_X$ and $F_Y$ are the CDFs of $X$ and $Y$. Measures monotonic relationship. Sample: $\hat{p}_s = \frac{\frac{1}{n}\sum r_i^{(x)} r_i^{(y)} - \bar{r}^{(x)}\bar{r}^{(y)}}{s_{r(x)} s_{r(y)}}$ Where $r_i^{(x)}$ is the rank of $x_i$ among $x_1, x_2, ..., x_n$

**Kendall's Tau**: For random variables $X$ and $Y$, $\tau = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0)$ Used for non-linear relationships. Sample: for each pair $(x_i, y_i)$ and $(x_j, y_j)$, count concordant pairs $C$ and discordant pairs $D$, then $\hat{\tau} = \frac{C-D}{\binom{n}{2}} = 4 \int \int F(x,y)f(x,y)dxdy - 1$

<div align="center">Copulas</div>

**Copula**: $C(t,s) : [0,1]^2 \to [0,1]$ to combine the marginal distributions $F_{X_i}(x)$ with correlations to form joint distribution $F_{X_1,X_2}(x_1,x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2))$ Note that $C(t,s) = F(F^{-1}(t), F^{-1}(s)) \iff F(x,y) = C(F(x), F(y))$ **Sklar's Theorem**: For any multivariate distribution $F$ with marginals $F_1, F_2, ..., F_n$, there exists a copula $C$ such that $F(x_1, x_2, ..., x_n) = C(F_1(x_1), F_2(x_2), ..., F_n(x_n))$. If $F_i$ are continuous, then $C$ is unique. Conversely, for any copula $C$ and marginals $F_i$, the function $F$ defined above is a multivariate distribution with marginals $F_i$.

**Cholesky Factorization**: For a positive definite matrix $\Sigma$, there exists a unique lower triangular matrix $L$ such that $\Sigma = LL^T$. Can generate correlated normals: $\Sigma = [\sigma_{ij}]$ then $a_{ij} = \frac{\sigma_{ij} - \sum_{k=1}^{j-1} a_{ik}a_{jk}}{a_{jj}}$ for $i \geq j$, $a_{ii} = \sqrt{\sigma_{ii} - \sum_{k=1}^{i-1} a_{ik}^2}$ Algo: For $j = 1$ to $n$, for $i = j$ to $n$, compute $v_i = \sigma_{ij} - \sum_{k=1}^{j-1} a_{ik}a_{jk}$, then set $a_{jj} = \sqrt{v_j}$, and for $i = j+1$ to $n$, set $a_{ij} = \frac{v_i}{a_{jj}}$.

**Multivar Normal Sim**: $X \sim N(\mu, \Sigma)$, $LL^T = \Sigma$ (Cholesky), $Z \sim N(0, I)$, then $X = LZ + \mu$

**Gaussian Copula**: Given correlation matrix $\Sigma$, the Gaussian copula is defined as $C(u_1, u_2, ..., u_n) = \Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2), ..., \Phi^{-1}(u_n))$ where $\Phi_\Sigma$ is the CDF of multivariate normal with covariance $\Sigma$ and $\Phi^{-1}$ is the inverse CDF of standard normal.

**Gaussian Copula Simulation**: $Z_n \sim N(0,1)$ iid $Z^* = LZ$, for $i, n$ $u_i = \Phi(z_i^*/\sigma_i)$, $x_i = F_i^{-1}(u_i)$ This gives correlated samples $x_i$ with marginals $F_i$ and correlation structure from $\Sigma$.

**Student t-Copula**: Same as gaussian but add $\chi^2(\nu)$ scaling in $F(Z_i^*/(\sigma_i\sqrt{W/\nu}))$ where $W \sim \chi^2(\nu)$ independent.

**Archimedian Copula**: Given a continuous, strictly decreasing function $\phi : [0,1] \to [0,\infty]$ with $\phi(1) = 0$, the Archimedian copula is defined as $C(u_1, u_2, ..., u_n) = \phi^{-1}(\phi(u_1) + \phi(u_2) + ... + \phi(u_n))$ where $\phi^{-1}$ is the pseudo-inverse of $\phi$. Common examples include Clayton, Gumbel, and Frank copulas.

<div align="center">Bootstrap</div>

**Bootstrap**: Using given data resample with replacement to create new datasets and estimate statistics. Algo: Take the bootstrap samples, calculate the statistic, order the statistics, and form confidence intervals.

**Symmetric CI**: $[\hat{\theta}_L, \hat{\theta}_U]$ where $L = \frac{\alpha}{2}N$, $U = (1 - \frac{\alpha}{2})N$

**Asymmetric CI**: $[2\hat{\theta} - \hat{\theta}_{(U)}, 2\hat{\theta} - \hat{\theta}_{(L)}]$.

**Bootstrap CLT** $\hat{\theta}^* - \hat{\theta}|\hat{\theta} \xrightarrow{d} \hat{\theta} - \theta_0|\theta_0$ as $n \to \infty$

**Bootstrap Residuals**: Fit reg line, calculate residuals, resample residuals with replacement, create new response variable, refit reg line, repeat B times to get bootstrap estimates.

<div align="center">Bayes</div>

**Bayes Theorem**: For events $A$ and $B$ with $P(B) > 0$, $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$. For continuous random variables, $f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}$.

**Prior** : Initial belief about parameter $\theta$ before observing data, denoted as $p(\theta)$. *Conjugate*: post same family as prior, *Elicited*: from expert knowledge, *Non-informative*: vague prior Jefferys prior invariant under transformation.

**Likelihood** : Probability of observed data given parameter $\theta$, denoted as $p(D|\theta)$. For i.i.d. data, $p(D|\theta) = \prod_{i=1}^{n} p(x_i|\theta)$.

**Posterior** : Updated belief about parameter $\theta$ after observing data, denoted as $p(\theta|D)$. Given by Bayes theorem: $p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$ where $p(D) = \int p(D|\theta)p(\theta)d\theta$.

**Bayesian Infrence**: Use posterior distribution to make inferences about $\theta$. Common summaries include posterior mean, median, mode, and credible intervals.

**Bayes Factor** $BF = \frac{P(D|M_1)}{P(D|M_2)}$ where $P(D|M_i) = \int P(D|\theta_i, M_i)P(\theta_i|M_i)d\theta_i$ is the marginal likelihood under model $M_i$. Used for model comparison. $1/10, 1/3, 1, 3, 10$ str M2, mod, weak, weak, mod, str M1 evidence.

**Bayes Example:** Data 13/16 pref A to B. Prior $Beta(.5, .5)$, Likelihood $Bin(16, \theta)$, Posterior $Beta(13+.5, 3+.5) = Beta(13.5, 3.5)$. $BF = \frac{\int_z^1 f(y|\theta)\pi(\theta)d\theta}{\int_0^z f(y|\theta)\pi(\theta)d\theta}$ where $z = $ critical value for H0: $\theta \geq z$ vs H1: $\theta < z$

**Linear Model Bayes:** $Y = X\beta + \epsilon$, $\epsilon \sim N(0, \sigma^2 I)$, observations $y|\beta, \sigma^2 \sim N(X\beta, \sigma^2 I)$, Prior: $\beta|\sigma^2 \sim N(\beta_0, \sigma^2 B_0)$, $\sigma^2 \sim \mathcal{G}(c, C_0)$, Posterior: $p(\beta, \sigma^2|y) \propto p(y|\beta, \sigma^2)p(\beta|\sigma^2)p(\sigma^2)$