MATH 559 FALL 2026
FORMULA SHEET

- **Complete-the-square** formulae for quadratic forms:
  - *Univariate case:*

$$A(x-a)^2 + B(x-b)^2 = (A+B)\left(x - \frac{Aa+Bb}{A+B}\right)^2 + \frac{AB}{A+B}(a-b)^2$$

  - *Multivariate case:*

$$(\mathbf{x}-\mathbf{a})^\top \mathbf{A}(\mathbf{x}-\mathbf{a}) + (\mathbf{x}-\mathbf{b})^\top \mathbf{B}(\mathbf{x}-\mathbf{b}) = (\mathbf{x}-\mathbf{m})^\top \mathbf{M}(\mathbf{x}-\mathbf{m}) + \mathbf{c}$$

    where

$$\mathbf{m} = (\mathbf{A}+\mathbf{B})^{-1}(\mathbf{A}\mathbf{a}+\mathbf{B}\mathbf{b}) \qquad \mathbf{M} = (\mathbf{A}+\mathbf{B})$$

    and

$$\mathbf{c} = \mathbf{a}^\top \mathbf{A}\mathbf{a} + \mathbf{b}^\top \mathbf{B}\mathbf{b} - \mathbf{m}^\top \mathbf{M}\mathbf{m}$$

- **Useful frequentist theory:** In the theory of maximum likelihood, we have under regularity conditions that the log-likelihood

$$\ell_n(\theta) = \sum_{i=1}^n \log f_Y(y_i;\theta) \qquad \theta \in \mathbb{R}^d$$

  admits a quadratic expansion around the maximizing value $\widehat{\theta}_n$

$$\ell_n(\theta) \doteq \ell_n(\widehat{\theta}_n) + \dot{\ell}_n(\widehat{\theta}_n)^\top(\widehat{\theta}_n - \theta) + \frac{1}{2}(\widehat{\theta}_n - \theta)^\top \ddot{\ell}_n(\widehat{\theta}_n)(\widehat{\theta}_n - \theta)$$

  where

$$\ddot{\ell}(\theta) = \sum_{i=1}^n \ddot{\ell}(y_i;\theta)$$

  with

$$\ddot{\ell}(y;\theta) = \frac{\partial^2 \ell(y;\theta)}{\partial\theta\partial\theta^\top} \qquad (d \times d).$$

  Noting that $\dot{\ell}_n(\widehat{\theta}_n) = 0$, we have that

$$\exp\{\ell_n(\theta)\} \doteq \exp\{\ell_n(\widehat{\theta}_n)\} \exp\left\{\frac{1}{2}(\widehat{\theta}_n - \theta)^\top \ddot{\ell}_n(\widehat{\theta}_n)(\widehat{\theta}_n - \theta)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}(\theta - \widehat{\theta}_n)^\top \{-\ddot{\ell}_n(\widehat{\theta}_n)\}(\theta - \widehat{\theta}_n)\right\}.$$

  Thus, for large n,

$$\pi_n(\theta) \doteq Normal_d\left(\widehat{\theta}_n, \left\{-\ddot{\ell}_n(\widehat{\theta}_n)\right\}^{-1}\right).$$

  and we have a Normal approximation to the posterior.

- **The Jeffreys Prior**

  The prior $\pi_0(\theta)$ for parameter θ in probability model $f_Y(y;\theta)$ determined by

  $$\pi_0(\theta) \propto \{|\mathcal{I}_\theta(\theta)|\}^{1/2}$$

  where $\mathcal{I}_\theta(\theta)$ is the *Fisher Information*,

  $$\mathcal{I}_\theta(\theta) = \mathbb{E}_Y\left[S(Y;\theta)S(Y;\theta)^\mathsf{T};\theta\right] = -\mathbb{E}_Y\left[\Psi(Y;\theta);\theta\right]$$

  and $|\mathcal{I}_\theta(\theta)|$ indicates the absolute value of the determinant of $\mathcal{I}_\theta(\theta)$ is termed the *Jeffreys* prior.

  - $S(y;\theta)$ is the $d \times 1$ vector *score function* with jth element

    $$S_j(y;\theta) = \frac{\partial}{\partial\theta_j}\log f_Y(y;\theta) \qquad j = 1,\ldots,d$$

  - $\Psi(Y;\theta)$ is the $d \times d$ matrix of second partial derivatives with (j,l)th element

    $$\frac{\partial^2}{\partial\theta_j\partial\theta_l}\log f_Y(y;\theta)$$

- **Kullback-Leibler Divergence**

  The *Kullback-Leibler* (KL) divergence is used when measuring the discrepancy between distributions. For two distributions with cdfs F_0, F_1

  $$KL(F_0,F_1) = \int \log\left\{\frac{dF_0(y)}{dF_1(y)}\right\} dF_0(y)$$

  which is defined when F_1 is absolutely continuous with respect to F_0, that is for the corresponding probability measures

  $$P_0(B) = 0 \quad \Longrightarrow \quad P_1(B) = 0$$

  for any set B.

  - Discrete case:

    $$KL(p_0,p_1) = \sum_y \log\left\{\frac{p_0(y)}{p_1(y)}\right\} p_0(y) = \mathbb{E}_{p_0}\left[\log\left\{\frac{p_0(Y)}{p_1(Y)}\right\}\right].$$

  - Continuous case:

    $$KL(f_0,f_1) = \int \log\left\{\frac{f_0(y)}{f_1(y)}\right\} f_0(y)\, dy = \mathbb{E}_{f_0}\left[\log\left\{\frac{f_0(Y)}{f_1(Y)}\right\}\right].$$

  1. $KL(F_0,F_1) \geq 0$;
  2. $KL(F_0,F_1) \neq KL(F_1,F_0)$ in general;
  3. $KL(F_0,F_1) = 0$ if and only if the two distributions are identical.

DISCRETE DISTRIBUTIONS							
	SUPPORT $\mathbb{X}$	PARAMETERS	MASS FUNCTION f_X	CDF F_X	$\mathbb{E}_X[X]$	$\text{Var}_X[X]$	MGF M_X
$Bernoulli(\theta)$	$\{0,1\}$	$\theta \in (0,1)$	$\theta^x(1-\theta)^{1-x}$		θ	$\theta(1-\theta)$	$1-\theta+\theta e^t$
$Binomial(n,\theta)$	$\{0,1,...,n\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{n}{x}\theta^x(1-\theta)^{n-x}$		$n\theta$	$n\theta(1-\theta)$	$(1-\theta+\theta e^t)^n$
$Poisson(\lambda)$	$\{0,1,2,...\}$	$\lambda \in \mathbb{R}^+$	$\frac{e^{-\lambda}\lambda^x}{x!}$		λ	λ	$\exp\{\lambda(e^t-1)\}$
$Geometric(\theta)$	$\{1,2,...\}$	$\theta \in (0,1)$	$(1-\theta)^{x-1}\theta$	$1-(1-\theta)^x$	$\frac{1}{\theta}$	$\frac{(1-\theta)}{\theta^2}$	$\frac{\theta e^t}{1-e^t(1-\theta)}$
$NegBinomial(n,\theta)$	$\{n,n+1,...\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{x-1}{n-1}\theta^n(1-\theta)^{x-n}$		$\frac{n}{\theta}$	$\frac{n(1-\theta)}{\theta^2}$	$\left(\frac{\theta e^t}{1-e^t(1-\theta)}\right)^n$
or	$\{0,1,2,...\}$	$n \in \mathbb{Z}^+, \theta \in (0,1)$	$\binom{n+x-1}{x}\theta^n(1-\theta)^x$		$\frac{n(1-\theta)}{\theta}$	$\frac{n(1-\theta)}{\theta^2}$	$\left(\frac{\theta}{1-e^t(1-\theta)}\right)^n$

For **CONTINUOUS** distributions (see over), define the **GAMMA FUNCTION** for $\alpha > 0$

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$$

and the LOCATION/SCALE transformation $Y = \mu + \sigma X$ gives

$$f_Y(y) = f_X\left(\frac{y-\mu}{\sigma}\right)\frac{1}{\sigma} \qquad F_Y(y) = F_X\left(\frac{y-\mu}{\sigma}\right) \qquad M_Y(t) = e^{\mu t}M_X(\sigma t) \qquad \mathbb{E}_Y[Y] = \mu+\sigma\mathbb{E}_X[X] \qquad \text{Var}_Y[Y] = \sigma^2\text{Var}_X[X]$$

CONTINUOUS DISTRIBUTIONS							
	SUPPORT $\mathbb{X}$	PARAMETERS	PDF f_X	CDF F_X	$\mathbb{E}_X[X]$	$\text{Var}_X[X]$	MGF M_X
$Uniform(\alpha, \beta)$ (standard model $\alpha = 0, \beta = 1$)	(α, β)	$\alpha < \beta \in \mathbb{R}$	$\frac{1}{\beta - \alpha}$	$\frac{x - \alpha}{\beta - \alpha}$	$\frac{(\alpha + \beta)}{2}$	$\frac{(\beta - \alpha)^2}{12}$	$\frac{e^{\beta t} - e^{\alpha t}}{t(\beta - \alpha)}$
$Exponential(\lambda)$ (standard model $\lambda = 1$)	$\mathbb{R}^+$	$\lambda \in \mathbb{R}^+$	$\lambda e^{-\lambda x}$	$1 - e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\left(\frac{\lambda}{\lambda - t}\right)$
$Gamma(\alpha, \beta)$ (standard model $\beta = 1$)	$\mathbb{R}^+$	$\alpha, \beta \in \mathbb{R}^+$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$		$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	$\left(\frac{\beta}{\beta - t}\right)^\alpha$
$Weibull(\alpha, \beta)$ (standard model $\beta = 1$)	$\mathbb{R}^+$	$\alpha, \beta \in \mathbb{R}^+$	$\alpha\beta x^{\alpha-1} e^{-\beta x^\alpha}$	$1 - e^{-\beta x^\alpha}$	$\frac{\Gamma\left(1 + \frac{1}{\alpha}\right)}{\beta^{1/\alpha}}$	$\frac{\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma\left(1 + \frac{1}{\alpha}\right)^2}{\beta^{2/\alpha}}$	
$Normal(\mu, \sigma^2)$ (standard model $\mu = 0, \sigma = 1$)	$\mathbb{R}$	$\mu \in \mathbb{R}$ $\sigma \in \mathbb{R}^+$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$		μ	σ^2	$\exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$
$Student(\nu)$	$\mathbb{R}$	$\nu \in \mathbb{R}^+$	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\left\{1 + \frac{x^2}{\nu}\right\}^{(\nu+1)/2}}$		0 (if $\nu > 1$)	$\frac{\nu}{\nu - 2}$ (if $\nu > 2$)	
$Pareto(\theta, \alpha)$	$\mathbb{R}^+$	$\theta, \alpha \in \mathbb{R}^+$	$\frac{\alpha\theta^\alpha}{(\theta + x)^{\alpha+1}}$	$1 - \left(\frac{\theta}{\theta + x}\right)^\alpha$	$\frac{\theta}{\alpha - 1}$ (if $\alpha > 1$)	$\frac{\alpha\theta^2}{(\alpha-1)(\alpha-2)}$ (if $\alpha > 2$)	
$Beta(\alpha, \beta)$	$(0, 1)$	$\alpha, \beta \in \mathbb{R}^+$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$		$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	

- **Multivariate distributions**

  The Multivariate Normal Distribution: The joint pdf of $\mathbf{X} = (X_1, \ldots, X_k)^\top$ takes the form

$$f_{X_1,\ldots,X_k}(x_1, \ldots, x_k) = \left(\frac{1}{2\pi}\right)^{k/2} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}$$

  where $\mathbf{x} = (x_1, \ldots, x_k)^\top$, $\boldsymbol{\mu}$ is a $k \times 1$ vector, and Σ is a symmetric, positive-definite $k \times k$ matrix. The distribution is obtained by taking a vector $\mathbf{Z} = (Z_1, \ldots, Z_k)^\top$ of independent standard Normal random variables with joint pdf

$$f_{Z_1,\ldots,Z_k}(z_1, \ldots, z_k) = \left(\frac{1}{2\pi}\right)^{k/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{k} z_i^2\right\} = \left(\frac{1}{2\pi}\right)^{k/2} \exp\left\{-\frac{1}{2}\mathbf{z}^\top \mathbf{z}\right\}$$

  and taking the linear transformation

$$\mathbf{X} = \mathbf{L}\mathbf{Z} + \boldsymbol{\mu}$$

  where $\mathbf{L}$ is the Cholesky factor of Σ, that is,

$$\Sigma = \mathbf{L}\mathbf{L}^\top.$$

  It can be shown that for any linear combination

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$$

  for constant matrix $\mathbf{A}$ and vector $\mathbf{b}$ (compatible in dimension) also has a multivariate normal distribution. The distribution of $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ can be deduced to be

$$\mathbf{Y} \sim Normal_k(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}^\top).$$

  All marginal and all conditional distributions derived from the multivariate normal are also multivariate normal. Let

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}.$$

  with subvectors of dimensions k_1 and $k_2 = k - k_1$ respectively. Let

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

  where Σ_{11} is $k_1 \times k_1$, Σ_{22} is $k_2 \times k_2$, $\Sigma_{21} = \Sigma_{12}^\top$.

  (i) **Marginal Distributions:** We have

$$\mathbf{X}_1 \sim Normal_{k_1}(\boldsymbol{\mu}_1, \Sigma_{11})$$

  (ii) **Conditional Distributions:** We have

$$\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1 \sim Normal_{k_2}\left(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)$$

FACTS ABOUT DISCRETE DISTRIBUTIONS

- Bernoulli/Binomial

$$Y_1, \ldots, Y_n \sim Bernoulli(p) \qquad \Longrightarrow \qquad Y = \sum_{i=1}^{n} Y_i \sim Binomial(n, p)$$

- Geometric/Negative Binomial

$$Y_1, \ldots, Y_n \sim Geometric(p) \qquad \Longrightarrow \qquad Y = \sum_{i=1}^{n} Y_i \sim NegBinomial(n, p)$$

- Binomial/Poisson

$$Y_n \sim Binomial(n, p) \longrightarrow Y \sim Poisson(\lambda)$$

  where $\lambda = np$ is held fixed and $n \longrightarrow \infty$.

- Negative Binomial/Poisson

$$Y_n \sim NegBinomial(n, p) \qquad X_n = Y_n - n \longrightarrow X \sim Poisson(\lambda)$$

  where $\lambda = n(1-p)$ is held fixed and $n \longrightarrow \infty$.

Sums of Independent Random Variables:

- Binomial

$$\left.\begin{array}{l} Y_1 \sim Binomial(m, p) \\ Y_2 \sim Binomial(n, p) \end{array}\right\} \qquad \Longrightarrow \qquad Y = Y_1 + Y_2 \sim Binomial(m + n, p)$$

- Negative Binomial

$$\left.\begin{array}{l} Y_1 \sim NegBinomial(m, p) \\ Y_2 \sim NegBinomial(n, p) \end{array}\right\} \qquad \Longrightarrow \qquad Y = Y_1 + Y_2 \sim NegBinomial(m + n, p)$$

- Poisson

$$\left.\begin{array}{l} Y_1 \sim Poisson(\lambda_1) \\ Y_2 \sim Poisson(\lambda_2) \end{array}\right\} \qquad \Longrightarrow \qquad Y = Y_1 + Y_2 \sim Poisson(\lambda_1 + \lambda_2)$$

FACTS ABOUT CONTINUOUS DISTRIBUTIONS

- Distributions on $\mathbb{R}^+$: Begin with $Y \sim Uniform(0,1)$:
  - $U = -\frac{1}{\beta}\log Y \sim Exponential(\beta)$, for $\beta > 0$.
  - $X = (\beta U/\lambda)^{1/\alpha} \sim Weibull(\alpha, \lambda)$, for $\alpha, \lambda > 0$.
  - If $X_1, \ldots, X_n \sim Exponential(\beta)$, independent, then $Z = \sum_{i=1}^{n} X_i \sim Gamma(n, \beta)$.
  - If $Y_1 \sim Gamma(\alpha_1, \beta)$ and $Y_2 \sim Gamma(\alpha_2, \beta)$ are independent, then
    $$S = Y_1 + Y_2 \sim Gamma(\alpha_1 + \alpha_2, \beta)$$
  - If $X \sim Gamma(\alpha, \beta)$ then
    $$Z = \frac{1}{X} \sim InverseGamma(\alpha, \beta)$$
    with
    $$f_Z(z) = \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{z}\right)^{\alpha+1}\exp\left\{-\frac{\beta}{z}\right\} \qquad z > 0.$$
- Distributions on $\mathbb{R}$: The Normal distribution and connections
  - Suppose $Y \sim Normal(0,1)$. Then $X = \mu + \sigma Y \sim Normal(\mu, \sigma^2)$.
  - Suppose $Y \sim Normal(0,1)$. Then $U = Y^2 \sim Gamma(1/2, 1/2) \equiv Chisquared(1)$.
  - If $Y_i \sim Gamma(\alpha_i/2, 1/2) \equiv Chisquared(\alpha_i)$ for $i = 1, \ldots, n$ are independent, then
    $$V = \sum_{i=1}^{n} Y_i \sim Gamma\left(\nu/2, 1/2\right) \equiv Chisquared(\nu)$$
    where
    $$\nu = \sum_{i=1}^{n} \alpha_i.$$
  - If $Y_1 \sim Normal(\mu_1, \sigma_1^2)$ and $Y_2 \sim Normal(\mu_2, \sigma_2^2)$ are independent, then
    $$Y = Y_1 + Y_2 \sim Normal(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$
- Distribution on $(0,1)$: The Beta distribution
  - If $Y_1 \sim Gamma(\alpha_1, \beta)$ and $Y_2 \sim Gamma(\alpha_2, \beta)$ are independent, then
    $$Y = \frac{Y_1}{Y_1 + Y_2} \sim Beta(\alpha_1, \alpha_2)$$
    This result follows by multivariate transformations.