Administrative Issues

- Homework 2 is due by September 25 at 11:59pm ET.
- Reminder: **No Handwritten Documents are permitted for any submission**

HW2 Guidance

1. **2D Density tasks**

   Marginal distribution for x:

   $$f_X(x) = \int_D f(x,y)dy$$

   Conditional distribution for y given x:

   $$f(y|x) = \frac{f(x,y)}{f_X(x)}$$

2. **Weibull Lifetimes**. From Wikipedia: The likelihood function (often simply called the likelihood) describes the joint probability of the observed data as a function of the parameters of the chosen statistical model.[1] For each specific parameter value θ in the parameter space, the likelihood function $p(\mathbf{X}|\theta)p(\mathbf{X}|\theta)$ therefore assigns a probabilistic prediction to the observed data $\mathbf{X}$. Since it is essentially the product of sampling densities, the likelihood generally encapsulates both the data-generating process as well as the missing-data mechanism that produced the observed sample.

   When you have more than one datapoint, the likelihood is calculated as follows:

   $$L = \prod_{i=1}^{n} f(x_i, y_i)$$

   where $f(x,y)$ is the pdf.

3. Memoryless property of exponential distributions is useful:

   If X is exponential with parameter $\lambda > 0$, then X is a **memoryless** random variable, that is $P(X > x + a|X > a) = P(X > x)$, for $a, x \geq 0$.

Other stuff

Likelihood	Prior	Posterior
$X_i \mid \theta \sim \mathcal{N}(\theta, \sigma^2)$	$\theta \sim \mathcal{N}(\mu, \tau^2)$	$\theta \mid \mathbf{X} \sim \mathcal{N}\left(\frac{\tau^2}{\tau^2+\sigma^2/n}\bar{X} + \frac{\sigma^2/n}{\tau^2+\sigma^2/n}\mu, \frac{\tau^2\sigma^2/n}{\tau^2+\sigma^2/n}\right)$
$X_i \mid \theta \sim \mathcal{B}in(m, \theta)$	$\theta \sim \mathcal{B}e(\alpha, \beta)$	$\theta \mid \mathbf{X} \sim \mathcal{B}e\left(\alpha + \sum_{i=1}^n X_i, \beta + mn - \sum_{i=1}^n X_i\right)$
$X_i \mid \theta \sim \mathcal{P}oi(\theta)$	$\theta \sim \mathcal{G}a(\alpha, \beta)$	$\theta \mid \mathbf{X} \sim \mathcal{G}a\left(\alpha + \sum_{i=1}^n X_i, \beta + n\right)$
$X_i \mid \theta \sim \mathcal{NB}(m, \theta)$	$\theta \sim \mathcal{B}e(\alpha, \beta)$	$\theta \mid \mathbf{X} \sim \mathcal{B}e\left(\alpha + mn, \beta + \sum_{i=1}^n X_i\right)$
$X_i \mid \theta \sim \mathcal{G}a(1/2, 1/(2\theta))$	$\theta \sim \mathcal{IG}(\alpha, \beta)$	$\theta \mid \mathbf{X} \sim \mathcal{IG}\left(\alpha + n/2, \beta + \frac{1}{2}\sum_{i=1}^n X_i\right)$
$X_i \mid \theta \sim \mathcal{U}(0, \theta)$	$\theta \sim \mathcal{P}a(\theta_0, \alpha)$	$\theta \mid \mathbf{X} \sim \mathcal{P}a\left(\max\{\theta_0, X_1, \ldots, X_n\}, \alpha + n\right)$
$X_i \mid \theta \sim \mathcal{N}(\mu, \theta)$	$\theta \sim \mathcal{IG}(\alpha, \beta)$	$\theta \mid \mathbf{X} \sim \mathcal{IG}\left(\alpha + n/2, \beta + \frac{1}{2}\sum_{i=1}^n (X_i - \mu)^2\right)$
$X_i \mid \theta \sim \mathcal{G}a(v, \theta)$	$\theta \sim \mathcal{G}a(\alpha, \beta)$	$\theta \mid \mathbf{X} \sim \mathcal{G}a\left(\alpha + nv, \beta + \sum_{i=1}^n X_i\right)$
$X_i \mid \theta \sim \mathcal{P}a(c, \theta)$	$\theta \sim \mathcal{G}a(\alpha, \beta)$	$\theta \mid \mathbf{X} \sim \mathcal{G}a\left(\alpha + n, \beta + \sum_{i=1}^n \log(X_i/c)\right)$

Note: To use this table correctly, please note that the left side is the likelihood for a single observation. It is **not** the joint likelihood, which needs to be computed and then multiplied with the prior to get the posterior.

For example: Likelihood is Normal $X_i \mid \theta \sim \mathcal{N}(\theta, \sigma^2)$ leads to a joint likelihood of

$$
\begin{aligned}
L(X_1, \ldots, X_n) &= \prod_{i=1}^n f(x_i, y_i) \\
&= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2n}} \exp\left(-\frac{(x_i - \theta)^2}{2\sigma^2}\right) \\
&= \frac{1}{\sigma^n (2n)^{1/2}} \exp\left(-\sum_{i=1}^n \frac{(x_i - \theta)^2}{2\sigma^2}\right)
\end{aligned}
$$

Example: Classical Statistics with MLE

Exponential Distribution with parameter λ:

$$f(t_i) = \lambda e^{-\lambda t_i}, \lambda > 0$$

The classical statistician would optimize for an unknown parameter given data points $1, \ldots, n$ using Maximum Likelihood Estimation. First we need the likelihood function:

$$L(\lambda; t_1, \ldots, t_n) = \prod_{i=1}^n \lambda e^{-\lambda t_i} = \lambda^n e^{-\lambda \sum_{i=1}^n t_i}$$

Due to the logarithm being strictly increasing, the maximum of $L()$ occurs at the same location as the maximum of $\log(L)$. So we take the logarithm first to make the algebra easier:

$$\begin{aligned}\ln L(\lambda; t_1, \ldots, t_n) &= \ln\left(\lambda^n e^{-\lambda \sum_{i=1}^n t_i}\right) \\ &= n\ln(\lambda) + \ln\left(e^{-\lambda \sum_{i=1}^n t_i}\right) \\ &= n\ln(\lambda) - \lambda \sum_{i=1}^n t_i\end{aligned}$$

Next we differentiate with respect to λ and set the result equal to 0:

$$\begin{aligned}\frac{d}{dx} \ln L(\lambda; t_1, \ldots, t_n) &= \frac{n}{\lambda} - \sum_{i=1}^n t_i \\ &= 0\end{aligned}$$

Rearranging and solving for λ yields

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n t_i}$$

Additional Examples - Posterior

1. Binomial likelihood and Beta prior

   Likelihood: $X \mid p \sim \text{Bin}(n, p)$; with associated pmf $f(x \mid p) = \left(\begin{array}{c} n \\ x \end{array}\right) p^x (1-p)^{n-x}$

   Prior: $p \sim Be(\alpha, \beta)$; with associated pdf $\pi(p) = \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}$

   We have

$$\begin{aligned}\pi(p \mid x) &\propto f(x \mid p)\pi(p) \\ &= \left(\begin{array}{c} n \\ x \end{array}\right) p^x (1-p)^{n-x} \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1} \\ &= \left(\begin{array}{c} n \\ x \end{array}\right) \frac{1}{B(\alpha,\beta)} p^{\alpha+x-1}(1-p)^{\beta+n-x-1} \\ &= C p^{\alpha+x-1}(1-p)^{\beta+n-x-1}\end{aligned}$$

   which is the kernel of a $\text{Beta}(\alpha + x, \beta + n - x)$ distribution.

2. Poisson likelihood and gamma prior

   Likelihood: $x \mid \lambda \sim \text{Poi}(\lambda)$; with pmf $f(x \mid \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$

   Prior: $\lambda \sim \text{Gamma}(\alpha, \beta)$; with pdf $\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$

$$
\begin{aligned}
\pi(\lambda \mid x) &\propto f(x \mid \lambda)\pi(\lambda) \\
&= \frac{\lambda^x e^{-\lambda}}{x!}\frac{\lambda^{\alpha-1}\beta^\alpha}{\Gamma(\alpha)}e^{-\beta\lambda} \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)x!}\lambda^x\lambda^{\alpha-1}e^{-\beta\lambda}e^{-\lambda} \\
&= C\lambda^{x+\alpha-1}e^{-(\beta+1)\lambda} \\
&= C\lambda^{a-1}e^{-b\lambda}
\end{aligned}
$$

which is the kernel of a Gamma$(x+\alpha, \beta+1)$ distribution.