STAT201A: Introduction to Probability at an Advanced Level (Fall 2024)
UC Berkeley

Problem Set 4

Due: 10:00pm, Tuesday, November 5, 2024 (via Gradescope)

1. **(Order statistics)** Let $X_1, \ldots, X_n$ be i.i.d. random variables with $\text{Exp}(\lambda)$ distribution, where $\lambda > 0$, and let $X_{(i)}$ be the order statistics for $i = 1, \ldots, n$.

   (a) Find the distribution of $X_{(1)}$.

   (b) Using the memoryless property, find the distribution of $X_{(i+1)} - X_{(i)}$ for $i = 1, \ldots, n-1$.

   (c) Use the previous item to show that each $X_{(i)}$ has the same distribution as a sum of i independent random variables.

   (d) Calculate the expectation and the variance of $X_{(i)}$ for $i = 1, \ldots, n$.

   (a) Since $X_{(1)} = \min_{i \in \{1,\ldots,n\}} X_i$, $\mathbb{P}(X_{(1)} > t) = \mathbb{P}(X_i > t)^n = e^{-n\lambda t}$. Hence $X_{(1)}$ is distributed as an exponential random variable with parameter $n\lambda$.

   (b) For each i, consider the $n-i$ random variables $Y_k = X_k - X_{(i)}|_{X_k > X_{(i)}}$. The key observation is that these random variables have exponential distributions, an application of the memoryless property gives
$$\mathbb{P}(Y_k > t) = \mathbb{P}(X_k - X_{(i)} > t | X_k > X_{(i)}) = \mathbb{P}(X_k > t).$$
   Moreover, $X_{(i+1)} - X_{(i)}$ corresponds to the minimum of the random variables Y_k and hence has exponential distribution with parameter $(n-i)\lambda$.

   (c) It follows from our prveious argument that we can write $X_i = \sum_{k=1}^{i} X_{(k)} - X_{(k-1)}$ where $X_{(0)} = 0$. Hence we can describe $X_{(i)} = \sum_{k=1}^{i} Z_k$ where Z_k is a collection of independent random variables, Z_k with exponential distribution with parameter $(n-k+1)\lambda$.

   (d) Finally, using our previous formula we have that
$$\mathbb{E}[X_{(i)}] = \sum_{k=1}^{i} \mathbb{E}[Z_k] = \sum_{k=1}^{i} \frac{1}{(n-k+1)\lambda}.$$
   Similarly for the variance,
$$\text{Var}[X_{(i)}] = \sum_{k=1}^{i} \text{Var}[Z_k] = \sum_{k=1}^{i} \frac{1}{(n-k+1)^2\lambda^2}.$$

2. **(Joint and conditional densities)** Let X, Y be two random variables with the following properties. Y has density function $f_Y(y) = 3y^2$ for $0 < y < 1$ and zero elsewhere. For $0 < y < 1$, given that $Y = y$, X has conditional density function $f_{X|Y}(x|y) = \frac{2x}{y^2}$ for $0 < x < y$ and zero elsewhere.

   (a) Find the joint density function $f_{X,Y}(x, y)$ of X, Y. Be precise about the values (x, y) for which your formula is valid. Check that the joint density function you find integrates to 1.

(b) Find the conditional density function of Y, given $X = x$. Be precise about the values of x and y for which the answer is valid. Identify the conditional distribution of Y by name.

(a) The joint density is given by

$$f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) = \frac{2x}{y^2}1_{0<x<y} \cdot 3y^2 1_{0<y<1} = 6x \cdot 1_{0<x<y<1}.$$

We have that

$$\int_{\mathbb{R}}\int_{\mathbb{R}} f_{X,Y}(x,y)\, dx\, dy = \int_0^1 \int_0^y 6x\, dx\, dy = 1.$$

(b) We first calculate the maringal density of X, $f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\, dy = \int_x^1 6x\, dx = 6x(1 - x) \cdot 1_{0<x<1}$. We can now calculate the conditional density of Y given X.

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{1}{1-x} \cdot 1_{x<y<1}.$$

We conclude that the conditional distribution of Y given X is $\mathcal{U}(X,1)$, uniform on the interval $(X,1)$.

3. **(Model selection)** Given data $x_1, \ldots, x_n$, consider the problem of selecting between the two models:

$$\text{Model One}: \ X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} N(0,1)$$

and

$$\text{Model Two}: \ X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} N(\mu,1) \text{ for an unknown } \mu.$$

To use probability to solve this problem, let us introduce an additional random variable Θ that has the Bernoulli distribution with parameter 0.5. Assume that the conditional distribution of $X_1, \ldots, X_n$ given $\Theta = \theta$ is given by the following

$$X_1, \ldots, X_n \mid \Theta = 0 \overset{\text{i.i.d}}{\sim} N(0,1)$$

and

$$X_1, \ldots, X_n \mid \mu, \Theta = 1 \overset{\text{i.i.d}}{\sim} N(\mu,1) \text{ and } \mu \mid \Theta = 1 \sim N\left(0, \tau^2\right).$$

Here τ is a parameter which you can treat as a fixed constant in this exercise.

(a) Using the formula

$$f_{X_1,\ldots,X_n|\Theta=1}(x_1,\ldots,x_n) = \int f_{X_1,\ldots,X_n|\mu,\Theta=1}(x_1,\ldots,x_n)\, f_{\mu|\Theta=1}(\mu) d\mu$$

prove that

$$f_{X_1,\ldots,X_n|\Theta=1}(x_1,\ldots,x_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sqrt{1+n\tau^2}} \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2}\right) \exp\left(\frac{n^2\tau^2\bar{x}^2}{2\left(1+n\tau^2\right)}\right),$$

where $\bar{x}$ is the mean of $x_1, \ldots, x_n$.

$$
\begin{aligned}
f_{X_1,\ldots,X_n|\Theta=1}(x_1,\ldots,x_n) &= \int f_{X_1,\ldots,X_n|\mu,\Theta=1}(x_1,\ldots,x_n)\, f_{\mu|\Theta=1}(\mu) d\mu \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right) \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{\mu^2}{2\tau^2}\right) d\mu \\
&= \frac{1}{\tau}\left(\frac{1}{\sqrt{2\pi}}\right)^{n+1} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 - 2\mu n\bar{x} + n\mu^2\right)\right) \exp\left(-\frac{\mu^2}{2\tau^2}\right) d\mu \\
&= \frac{1}{\tau}\left(\frac{1}{\sqrt{2\pi}}\right)^{n+1} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\left(n+\frac{1}{\tau^2}\right)\mu^2 - 2n\mu\bar{x}\right)\right) d\mu \\
&= \frac{1}{\tau}\left(\frac{1}{\sqrt{2\pi}}\right)^{n+1} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right) \sqrt{\frac{2\pi}{n+\frac{1}{\tau^2}}} \exp\left(\frac{n^2\tau^2\bar{x}^2}{2(1+n\tau^2)}\right) \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sqrt{1+n\tau^2}} \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2}\right) \exp\left(\frac{n^2\tau^2\bar{x}^2}{2(1+n\tau^2)}\right).
\end{aligned}
$$

(b) Calculate the conditional distribution of Θ given $X_1 = x_1, \ldots, X_n = x_n$.
From Model One, we know

$$
f_{X_1,\ldots,X_n|\Theta=0}(x_1,\ldots,x_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2}\right).
$$

By Bayes' theorem,

$$
\begin{aligned}
&\mathbb{P}\left(\Theta = 1 \mid X_1 = x_1, \ldots, X_n = x_n\right) \\
&= \frac{\mathbb{P}(\Theta=1) f_{X_1,\ldots,X_n|\Theta=1}(x_1,\ldots,x_n)}{\mathbb{P}(\Theta=0) f_{X_1,\ldots,X_n|\Theta=0}(x_1,\ldots,x_n) + \mathbb{P}(\Theta=1) f_{X_1,\ldots,X_n|\Theta=1}(x_1,\ldots,x_n)} \\
&= \frac{f_{X_1,\ldots,X_n|\Theta=1}(x_1,\ldots,x_n)}{f_{X_1,\ldots,X_n|\Theta=0}(x_1,\ldots,x_n) + f_{X_1,\ldots,X_n|\Theta=1}(x_1,\ldots,x_n)} \\
&= \frac{\frac{1}{\sqrt{1+n\tau^2}} \exp\left(\frac{n^2\tau^2\bar{x}^2}{2(1+n\tau^2)}\right)}{1 + \frac{1}{\sqrt{1+n\tau^2}} \exp\left(\frac{n^2\tau^2\bar{x}^2}{2(1+n\tau^2)}\right)}
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathbb{P}\left(\Theta = 0 \mid X_1 = x_1, \ldots, X_n = x_n\right) &= 1 - \mathbb{P}\left(\Theta = 1 \mid X_1 = x_1, \ldots, X_n = x_n\right) \\
&= \frac{1}{1 + \frac{1}{\sqrt{1+n\tau^2}} \exp\left(\frac{n^2\tau^2\bar{x}^2}{2(1+n\tau^2)}\right)}
\end{aligned}
$$

(c) Intuitively, we would prefer Model Two over Model One when $\bar{x}$ is far from zero. Is this intuition reflected in your conditional distribution from the previous part?

Yes. When $\bar{x}$ is close to zero, the exponential term $\exp\left(\frac{n^2\tau^2\bar{x}^2}{2(1+n\tau^2)}\right)$ is approximately 1, and the square root term $\frac{1}{\sqrt{1+n\tau^2}}$ is small for large n, so $(\Theta = 1)$ remains small, meaning we favor Model One. As $\bar{x}$ moves away from zero, the exponential term grows rapidly, dominating the expression, so $\mathbb{P}(\Theta = 1)$ increases and close to 1, favoring Model Two.

4. **(Gamma-Poisson)** Consider random variables $\Theta, X_1, \ldots, X_n$ such that

$$\Theta \sim \text{Gamma}(\alpha, \lambda) \quad \text{and} \quad X_1, \ldots, X_n \mid \Theta = \theta \overset{\text{i.i.d.}}{\sim} \text{Poisson}(\theta)$$

(a) Find the conditional distribution of Θ given $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.
The conditional distribution of Θ given $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ is

$$\begin{aligned}
f_{\Theta|X_1=x_1,X_2=x_2,\ldots,X_n=x_n}(\theta) &\propto f_{X_1,X_2,\ldots,X_n|\Theta=\theta}(x_1, x_2, \ldots, x_n) f_\Theta(\theta) \\
&\propto \left(\prod_{i=1}^n f_{X_i|\Theta=\theta}(x_i)\right) f_\Theta(\theta) \\
&= \left(\prod_{i=1}^n \frac{\theta^{x_i} e^{-\theta}}{x_i!} 1\{x_i \in \mathbb{N}_0\}\right) \left(\frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta} 1\{\theta > 0\}\right) \\
&= \frac{\theta^{\sum x_i} e^{-n\theta}}{\prod_{i=1}^n x_i!} 1\{x_i \in \mathbb{N}_0\} \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta} 1\{\theta > 0\} \\
&\propto \theta^{\alpha-1+\sum x_i} e^{-(n+\lambda)\theta} 1\{x_i \in \mathbb{N}_0\} 1\{\theta > 0\}.
\end{aligned}$$

The above is in the form of a Gamma $(\alpha + \sum x_i, n + \lambda)$ distribution. Therefore, the exact distribution of Θ given $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ is

$$f_{\Theta|X_1=x_1,X_2=x_2,\ldots,X_n=x_n}(\theta) = \frac{(n+\lambda)^{\alpha+\sum x_i}}{\Gamma(\alpha + \sum x_i)} \theta^{\alpha+\sum x_i - 1} e^{-(n+\lambda)\theta} 1\{\theta > 0\},$$

where $x_i \in \mathbb{N}_0$.

(b) Find $\mathbb{E}[\Theta \mid X_1 = x_1, \ldots, X_n = x_n]$.
We know that for $Y \sim \text{Gamma}(\alpha, \beta)$, $\mathbb{E}(Y) = \frac{\alpha}{\beta}$. Since $\Theta \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \sim \text{Gamma}(\alpha + \sum x_i, n + \lambda)$, we have

$$\mathbb{E}[\Theta \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n] = \frac{\alpha + \sum x_i}{n + \lambda}.$$

(c) Write $\mathbb{E}[\Theta \mid X_1 = x_1, \ldots, X_n = x_n]$ as a weighted linear combination of $(x_1 + \cdots + x_n)/n$ and the mean of the marginal distribution (i.e., prior mean) of Θ and argue that the weight of the prior mean goes to zero as $n \to \infty$.
We can express $\mathbb{E}[\Theta \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n]$ as a weighted linear combination of the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ and the prior mean $\frac{\alpha}{\lambda}$. Specifically, we have

$$\mathbb{E}[\Theta \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n] = \frac{\alpha + \sum_{i=1}^n x_i}{n + \lambda} = \frac{\lambda}{n+\lambda} \cdot \frac{\alpha}{\lambda} + \frac{n}{n+\lambda} \cdot \bar{x}.$$

As $n \to \infty$, the weight on the prior mean, $\frac{\lambda}{n+\lambda}$, tends to 0, meaning the prior mean becomes less influential. Conversely, the weight on the sample mean, $\frac{n}{n+\lambda}$, tends to 1, meaning the sample mean dominates as n increases. Thus, as $n \to \infty$, the conditional expectation of Θ approaches the sample mean $\bar{x}$, which aligns with the intuition that with more data, the influence of the prior diminishes, and the posterior is dominated by the data.

5. **(Law of total expectation)** Let the joint probability mass function (p.m.f.) of (X, Y) be

$$p_{X,Y}(k,n) = \begin{cases} \dfrac{1}{n+1}\left(1 - \dfrac{1}{n+1}\right)^{k-1}\dfrac{1}{2^n}, & \text{for } 1 \le n < \infty \text{ and } 1 \le k < \infty, \\ 0, & \text{else.} \end{cases}$$

(a) Find the p.m.f. $p_Y(n)$ of Y and the conditional p.m.f $p_{X|Y}(k|n)$.

(b) Calculate $\mathbb{E}[Y]$.

(c) Find the conditional expectation $\mathbb{E}[X|Y]$.

(d) Use parts (a) and (c) to calculate $\mathbb{E}[X]$.

(a) We start with a calculation.

$$\begin{aligned} p_Y(n) &= \sum_k p_{X,Y}(k,n) \\ &= \sum_{k=1}^{\infty} \frac{1}{n+1}(1 - \frac{1}{n+1})^{k-1}\frac{1}{2^n} \\ &= \frac{1}{(n+1)2^n}\sum_{k=1}^{\infty}(1 - \frac{1}{n+1})^{k-1} \\ &= \frac{1}{(n+1)2^n}(n+1) = \frac{1}{2^n}. \end{aligned}$$

Now we have that $p_{X|Y}(k|n) = \dfrac{p_{X,Y}(k,n)}{p_Y(n)} = \dfrac{1}{n+1}(1 - \dfrac{1}{n+1})^{k-1}$. We are able to recognize this distributions.

$$Y \sim \text{Geom}(1/2) \quad \text{and} \quad X|_{Y=n} \sim \text{Geom}(1/(n+1)).$$

(b) We automatically conclude from (a) that $\mathbb{E}[Y] = 2$.

(c) Since $X|_{Y=n} \sim \text{Geom}(1/(n+1))$ we automatically conclude that $\mathbb{E}[X|Y = n] = n + 1$. It follows that $\mathbb{E}[X|Y] = Y + 1$.

(d) We finally have that $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[Y + 1] = \mathbb{E}[Y] + 1 = 3$.

6. **(Expected number of coin tosses)** Consider a sequence of coin tosses.

(a) On average, how many tosses of a fair coin does it take to see two heads in a row?

(b) How many tosses on average to see the sequence HTH for the first time?

(c) How does our answer changes if we have an unfair coin?

(a) Let X be the random variable describing the number of tosses needed to see tho heads in a row. Let A be the even that the first toss is tails, B the event that the first two tosses are heads and C the event that the first toss is head and the second is tails. Observe that the following equation is satisfied

$$\mathbb{E}[X] = \mathbb{E}[X|A]\mathbb{P}(A) + \mathbb{E}[X|B]\mathbb{P}(B) + \mathbb{E}[X|C]\mathbb{P}(C).$$

Now the key observation is that $\mathbb{E}[X|A] = \mathbb{E}[X] + 1$, $\mathbb{E}[X|B] = 2$ and $\mathbb{E}[X|C] = \mathbb{E}[X] + 2$. We can now solve the equation

$$\mathbb{E}[X] = \frac{\mathbb{E}[X] + 1}{2} + \frac{2}{4} + \frac{\mathbb{E}[X] + 2}{4}.$$

We get $\mathbb{E}[X] = 6$.

(b) We repeat the previous argument. Let x represents the expected number of tosses to get HTH, y the expected number of tosses to get HTH given that our last toss is H and z the expected number of tosses to get HTH given that our last toss is HT. We then obtain the following system of equations. $a = \frac{a+1}{2} + \frac{b+1}{2}$. $b = \frac{b+1}{2} + \frac{c+1}{2}$ and finally $c = \frac{1}{2} + \frac{a+1}{2}$. Solving the system of equations gives $a = 10$.

(c) This is completly analogous to part (a) and (b) only that the probability to get H is now p. Solving the equations we get that the expected number of tosses to get HH is $\frac{1+p}{p^2}$ while the expected number of tosses to get HTH is $\frac{1+p-p^2}{p^2(1-p)}$.