MA 413 - Statistics for Data Science

Solutions to Exercise 4

1. We have that $I(X,Y)=0$, that is
$$\int\int f_{X,Y}(X,Y)\log\left[\frac{f_X(X)f_Y(Y)}{f_{X,Y}(X,Y)}\right]\,dx\,dy=0$$
The conclusion would have been straightforward had it been that log was a non-negative function. But since this is not the case, we proceed as follows: Notice that
$$e^x=e\cdot e^{x-1}=e\cdot\left[1+(x-1)+\frac{(x-1)^2}{2!}+\cdots\right]\geq e\cdot(1+x-1)=ex$$
So, for all x we have $e^x\geq ex$ and thus $\log x\leq x-1$ for all $x\geq 0$ with equality only if $x=1$ $(*)$. Now since $0\leq x-\log x-1$ for $x\geq 0$ we have
$$0\leq\frac{f_X(x)f_Y(y)}{f_{X,Y}(x,y)}-\log\left[\frac{f_X(x)f_Y(y)}{f_{X,Y}(x,y)}\right]-1=g(x,y)\ \text{(say)}$$
Now, notice further that
$$\int\int f_{X,Y}(x,y)g(x,y)\,dx\,dy=1+I(X,Y)-1=I(X,Y)$$
Clearly, $I(X,Y)=0$ if and only if $g(x,y)=0$ for all x,y. But by $(*)$, this is only if
$$\frac{f_X(x)f_Y(y)}{f_{X,Y}(x,y)}=1$$
and the conclusion follows.
Remark. The above solution is elementary but unintuitive. A more principled approach is to use Jensen's inequality which states that for a convex function f and any random variable X we have: $\mathbb{E}[f(X)]\leq f(\mathbb{E}[X])$ with equality only if f is linear on the range of X. Taking $f(x)=-\log x$ and X to be $\frac{f_X(X)f_Y(Y)}{f_{X,Y}(X,Y)}$ one can show that $I(X,Y)\geq 0$ and using the condition for equality, that $I(X,Y)=0$ if and only if X and Y are independent.

2. For $X\sim\text{Pois}(\lambda)$, we have for $k\geq 0$,
$$\begin{aligned}f_X(k)&=e^{-\lambda}\frac{\lambda^k}{k!}\\&=\exp\left[(\log\lambda)k-\lambda-\log k!\right]\end{aligned}$$
Usual parametrization in terms of Poisson mean implies $\theta=\lambda$. So we can write, $\eta(\theta)=\log\theta$, $T(k)=k$, $d(\theta)=\theta$ and $S(k)=-\log k!$.

Remark. As you can see, the parametrization is not unique. One might as well write, $\theta=\lambda$, $\eta(\theta)=\log\theta$, $T(k)=k-1$, $d(\theta)=\theta-\log\theta$ and $S(k)=-\log k!$.

3. For $X \sim \text{Gamma}(\alpha, \beta)$, we have for $x \geq 0$,

$$\begin{aligned} f_X(x) &= \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \\ &= \exp\left[(\alpha - 1)\log x - \beta x + \alpha \log \beta - \log \Gamma(\alpha)\right] \\ &= \exp\left[\alpha \log x + \beta(-x) + \alpha \log \beta - \log \Gamma(\alpha) - \log x\right] \end{aligned}$$

Usual parametrization in terms of Gamma parameters means $\theta_1 = \alpha$ and $\theta_2 = \beta$. So we write,

$$\begin{aligned} \eta_1(\theta_1, \theta_2) &= \theta_1 & T_1(x) &= \log x \\ \eta_2(\theta_1, \theta_2) &= \theta_2 & T_2(x) &= -x \\ d(\theta_1, \theta_2) &= -\theta_1 \log \theta_2 + \log \Gamma(\theta_1) & S(x) &= -\log x \end{aligned}$$

Remark. Again the parametrization is not unique. In fact, I have tried to make η_1 and η_2 as simple as possible, which is completely unnecessary!

4. Let $X_j \sim \text{Geom}(p)$ for $j = 1, \ldots, m$ be independent random variables. We are interested in finding the distribution of the sum $X = \sum_{j=1}^m X_j$. So, we calculate its moment generating function as follows,

$$\begin{aligned} M_X(t) &= \mathbb{E}\left[e^{tX}\right] \\ &= \mathbb{E}\left[e^{tX_1} \cdot e^{tX_2} \cdots e^{tX_m}\right] \\ &= \mathbb{E}\left[e^{tX_1}\right] \cdot \mathbb{E}\left[e^{tX_2}\right] \cdots \mathbb{E}\left[e^{tX_m}\right] \\ &= \left[\frac{p}{1-(1-p)e^t}\right] \cdots \left[\frac{p}{1-(1-p)e^t}\right] = \left[\frac{p}{1-(1-p)e^t}\right]^m \end{aligned}$$

This is same as the moment generating function of a negative binomial random variable with the parameters m and p. Therefore, $X \sim \text{NegBin}(m, p)$.

Note. Here is how you compute the moment generating function for $X \sim \text{Geom}(p)$,

$$\mathbb{E}\left[e^{tX}\right] = \sum_{k=0}^{\infty} e^{tk}(1-p)^k p = p \cdot \sum_{k=0}^{\infty} \left[(1-p)e^t\right]^k = \frac{p}{1-(1-p)e^t}$$

Remark. Be assured that adequate information will be provided to you if problems involving unfamiliar distributions such as the negative binomial appear in the final exam.

5. Let $X_j \sim \text{Gamma}(\alpha_j, \beta)$ for $j = 1, \ldots, k$ be independent random variables. As before we are interested in finding the distribution of the sum $X = \sum_{j=1}^m X_j$ and so we calculate its moment generating function,

$$\begin{aligned} M_X(t) &= M_{X_1}(t) \cdots M_{X_k}(t) \\ &= \left[\frac{\beta}{\beta - t}\right]^{\alpha_1} \cdots \left[\frac{\beta}{\beta - t}\right]^{\alpha_k} = \left[\frac{\beta}{\beta - t}\right]^{\sum_{j=1}^k \alpha_j} \end{aligned}$$

which is same as the moment generating function of a Gamma random variable with the parameters given by $\alpha = \sum_{j=1}^{k} \alpha_j$ and β. Therefore, $X \sim \text{Gamma}(\sum_{j=1}^{k} \alpha_j, \beta)$.

Note. This is how you compute the moment generating function for $X \sim \text{Gamma}(\alpha, \beta)$,

$$\begin{aligned}
M_X(t) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{tx} \cdot x^{\alpha-1} e^{-\beta x} \, dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\beta-t)x} \, dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\beta-t)^\alpha} = \left[\frac{\beta}{\beta-t}\right]^\alpha
\end{aligned}$$

for $t < \beta$ and ∞, otherwise. Here we evaluated the Gamma integral using the knowledge that the density of the Gamma distribution integrates to one over the real line.

6. Let $X \sim \text{Pois}(\lambda)$. Then its p.m.f. is given by $f_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ for $k \geq 0$. So we calculate,

$$M_X(t) = \sum_{k=0}^{\infty} e^{tk} \cdot e^{-\lambda} \tfrac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \tfrac{(\lambda e^t)^k}{k!} = e^{-\lambda} e^{\lambda e^t}$$

Therefore, $M_X(t) = e^{\lambda e^t - \lambda}$.

7. Let $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$ for $j = 1, \ldots, k$. Since, $M_{X_j}(t) = e^{\mu_j t + \sigma_j^2 t^2/2}$ we have,

$$M_X(t) = M_{X_1}(t) \cdots M_{X_k}(t) = \exp\left[\left(\sum_{j=1}^{k} \mu_j\right) t + \frac{1}{2}\left(\sum_{j=1}^{k} \sigma_j^2\right) t^2\right]$$

which is the moment generating function of a normal random variable with mean $\mu = \sum_{j=1}^{k} \mu_j$ and variance $\sigma^2 = \sum_{j=1}^{k} \sigma_j^2$. The conclusion follows.

8. Be definition, we have

$$\begin{aligned}
\mathbb{E}_{Y|X}\left[Y|X\right] &= \int y \cdot f_{Y|X}(y|x) \, dy = \frac{1}{f_X(x)} \int y \cdot f_{X,Y}(x,y) \, dy \\
\mathbb{E}_{Y|X}\left[Y^2|X\right] &= \int y^2 \cdot f_{Y|X}(y|x) \, dy = \frac{1}{f_X(x)} \int y^2 \cdot f_{X,Y}(x,y) \, dy
\end{aligned}$$

Taking the expectation with respect to X yields,

$$\begin{aligned}
\mathbb{E}_X\left[\mathbb{E}_{Y|X}\left[Y^2|X\right]\right] &= \int \left[\frac{1}{f_X(x)} \int y^2 \cdot f_{X,Y}(x,y) \, dy\right] \cdot f_X(x) \, dx \\
&= \int \int y^2 \cdot f_{X,Y}(x,y) \, dy \, dx = \mathbb{E}_Y\left[Y^2\right] \\
\mathbb{E}_X\left[\mathbb{E}_{Y|X}[Y|X]\right] &= \int \left[\frac{1}{f_X(x)} \int y \cdot f_{X,Y}(x,y) \, dy\right] \cdot f_X(x) \, dx \\
&= \int \int y \cdot f_{X,Y}(x,y) \, dy \, dx = \mathbb{E}_Y\left[Y\right]
\end{aligned}$$

We begin by simplifying the first term,

$$
\begin{aligned}
\mathrm{Var}_{Y|X}(Y|X) &= \mathbb{E}_{Y|X}\left[Y^2|X\right] - \left[\mathbb{E}_{Y|X}\left[Y|X\right]\right]^2 \\
\mathbb{E}_X\left[\mathrm{Var}_{Y|X}(Y|X)\right] &= \mathbb{E}_X\left[\mathbb{E}_{Y|X}\left[Y^2|X\right]\right] - \mathbb{E}_X\left[\mathbb{E}_{Y|X}\left[Y|X\right]^2\right] \\
\mathbb{E}_X\left[\mathrm{Var}_{Y|X}(Y|X)\right] &\overset{1}{=} \mathbb{E}_Y\left[Y^2\right] - \mathbb{E}_X\left[\mathbb{E}_{Y|X}\left[Y|X\right]^2\right]
\end{aligned}
$$

Now we treat the second term,

$$
\begin{aligned}
\mathrm{Var}_X\left[\mathbb{E}_{Y|X}[Y|X]\right] &= \mathbb{E}_X\left[\mathbb{E}_{Y|X}\left[Y|X\right]^2\right] - \left[\mathbb{E}_X\left[\mathbb{E}_{Y|X}[Y|X]\right]\right]^2 \\
\mathrm{Var}_X\left[\mathbb{E}_{Y|X}[Y|X]\right] &\overset{2}{=} \mathbb{E}_X\left[\mathbb{E}_{Y|X}\left[Y|X\right]^2\right] - \left[\mathbb{E}_Y\left[Y\right]\right]^2
\end{aligned}
$$

From equations (1) and (2) it follows that,

$$
\mathrm{Var}_Y\left[Y\right] = \mathbb{E}_X\left[\mathrm{Var}_{Y|X}(Y|X)\right] + \mathrm{Var}_X\left[\mathbb{E}_{Y|X}[Y|X]\right]
$$

The trick is to simplify the more complicated terms simultaneously, while looking out for any terms which may cancel each other out. Proving this for covariances is very similar. One only needs to replace some of the square terms with product terms.