Lab 6

Multivariate RV's: Special Topics

6.1. Convolutions

Definition 6.1.1. The *convolution* of two functions f and g, where $\int_{\mathbb{R}} |f(x)|dx, \int_{\mathbb{R}} |g(x)|dx < \infty$, is defined as

$$(f * g)(t) = \int_{\mathbb{R}} f(x)g(t-x)dx$$

Some properties of convolution include

Commutativity: $f * g = g * f$

Associativity: $f * (g * h) = (f * g) * h$

Distributivity: $f * (g + h) = (f * g) + (f * h)$

Associativity with scalar multiplication: $a(f * g) = (af) * g$

For those more math-savy, these properties imply that the convolution on the space of integrable functions is a commutative algebra without identity. Convolutions are important in probability theory because, in finding, say, the density of $X_1 + X_2$, where each is independent and follows densities f_{X_1} and f_{X_2}, respectively, we have

$$F_{X_1+X_2}(t) = \int_{-\infty}^{\infty} f_{X_1}(x)F_{X_2}(t-x)dx$$
$$\implies f_{X_1+X_2}(t) = \int_{-\infty}^{\infty} f_{X_1}(x)f_{X_2}(t-x)dx = (f_{X_1} * f_{X_2})(t)$$

Going through the actual convolution can be algebraically painful, which is why in practice one often relies on...

Theorem 6.1.1 (Convolution Theorem)**.** $\mathcal{F}\{f * g\} = \mathcal{F}\{f\} \cdot \mathcal{F}\{g\}$, where $\mathcal{F}$ is either the Fourier or two-sided Laplace transform.

In problems where X_1 and X_2 are iid, say with common f_X, we can often recognize the common $f_{X_1+X_2}$ such that $\mathcal{F}\{f_{X_1+X_2}\} = \mathcal{F}\{f_{X_1}\} \cdot \mathcal{F}\{f_{X_2}\}$, and hence applying the inverse transform $\mathcal{F}^{-1}$, we arrive with $f_{X_1} * f_{X_2} = f_{X_1+X_2}$. We do this all the time when we multiply the CF's or MGF's of iid random variables. These functions correspond to the Fourier and two-sided LaPlace transform, respectively, and hence the basis for determining the resulting distribution function implicitly relies on the Convolution Theorem. Of course, we can generalize

$$\begin{aligned}
X_1, \cdots, X_n \overset{\text{iid}}{\sim} f_X &\implies X_1 + \cdots + X_n \sim f_{X_1} * \cdots * f_{X_n} \\
&\iff X_1 + \cdots + X_n \overset{\text{CF}}{\sim} \mathcal{F}\{f_{X_1} * \cdots * f_{X_n}\} \\
&\iff X_1 + \cdots + X_n \overset{\text{CF}}{\sim} \mathcal{F}\{f_{X_1}\} \times \cdots \times \mathcal{F}\{f_{X_n}\} \quad \text{Convolution Theorem} \\
&\iff X_1 + \cdots + X_n \overset{\text{CF}}{\sim} \mathcal{F}\{f_{X_1 + \cdots + X_n}\} \quad \text{``Recognize''} \\
&\iff X_1 + \cdots + X_n \sim f_{X_1 + \cdots + X_n}
\end{aligned}$$

The "recognize" step is not always guaranteed, and fits very specifically to certain distributions. We present a few of these distributions below:

f_{X_i}	$f_{\sum X_i}$ (X_i independent)
$\text{Ber}(p)$	$\text{Bin}(n, p)$
$\text{Bin}(n_i, p)$	$\text{Bin}(\sum_i n_i, p)$
$\text{Geo}(p)$	$\text{NBin}(n, p)$
$\text{Exp}(\lambda)$	$\text{Gam}(n, \lambda)$ – Note: λ is rate parameter
$\text{Gam}(n_i, \beta)$	$\text{Gam}(\sum_i n_i, \beta)$
$\text{Pois}(\lambda_i)$	$\text{Pois}(\sum_i \lambda_i)$
$\chi^2(1)$	$\chi^2(n)$
$\chi^2(n_i)$	$\chi^2(\sum_i n_i)$
$\text{N}(\mu_i, \sigma_i^2)$	$\text{N}(\sum_i \mu_i, \sum_i \sigma_i^2)$

Please take due diligence in verifying the table above; such "verification" could indeed be an exam question!

6.2. Iterated Moments

One of the most useful techniques in probability and statistics is

Theorem 6.2.1. We have

$$\begin{aligned}
\mathbb{E}X &= \mathbb{E}[\mathbb{E}[X|Y]] && \text{Law of Total Expectation} \\
\text{Var}(X) &= \mathbb{E}\text{Var}(X|Y) + \text{Var}(\mathbb{E}[X|Y]) && \text{Law of Total Variance}
\end{aligned}$$

Corollary 6.2.1. Suppose $X_1, \cdots, X_n$ are iid, and N is independent of the X_i's. If $S_N = \sum_{i=1}^N X_i$, we have

$$\begin{aligned}
\mathbb{E}S_N &= (\mathbb{E}N)(\mathbb{E}X_1) && \text{Wald's equation, assuming } \mathbb{E}|X_i|, \mathbb{E}|N| < \infty \\
\text{Var}(S_N) &= (\mathbb{E}N)(\text{Var}(X_1)) + (\mathbb{E}X_1)^2(\text{Var}(N)) && \text{Blackwell-Girshick equation,} \\
& && \text{assuming } \mathbb{E}|X_i|^2, \mathbb{E}|N|^2 < \infty
\end{aligned}$$

6.3. Worked Examples

Example 6.3.1. We want to model the number of lung cancer cases (Y) as a function of covariates $\mathbf{X} = (X_1, \cdots, X_n)$ (i.e. PM_{10}, radon concentration, etc). Suppose the true process follows the *Poisson model*

$$\log(\mathbb{E}[Y|\mathbf{X}]) = \alpha + \boldsymbol{\beta}^\intercal \mathbf{X}$$

Assuming X_i are mutually independent from each other, determine $\mathbb{E}Y$.

Solution. We have

$$\mathbb{E}Y = \mathbb{E}[\mathbb{E}[Y|\mathbf{X}]] = e^\alpha \mathbb{E} e^{\boldsymbol{\beta}^\intercal \mathbf{X}} = e^\alpha \prod_{i=1}^{n} M_{X_i}(\beta_i)$$

□

Example 6.3.2. An elevator services N floors above the ground floor. If $R \sim \text{Pois}(\lambda)$ riders enter the elevator on the ground floor, and each rider is equally likely to get off at any of the N possible floors, compute the expected number of stops S.

Solution. Let $S_i = 1$ if the elevator stops on the ith floor, and $= 0$ if not. Then $S = \sum_{i=1}^N S_i$. We may compute

$$\mathbb{E}[S_i|R] = 1 - \left(1 - \frac{1}{N}\right)^R$$

Hence $\mathbb{E}[S|R] = N(1 - \left(1 - \frac{1}{N}\right)^R)$. Applying iterated expectations,

$$\mathbb{E}S = N \sum_{r=0}^{\infty} \left(1 - \left(1 - \frac{1}{N}\right)^r\right) \frac{e^{-\lambda}\lambda^r}{r!} = N(1 - e^{-\lambda/N})$$

□

Example 6.3.3. Let $X_{k-1}|X_k \sim \text{Unif}(0, X_k)$ for $2 \le k \le n$ and $X_n \sim \text{Unif}(0, 1)$. Find the marginal density of X_1.

Solution. Denote $\mathbf{X}_{i:j} = (X_i, X_{i+1}, \cdots, X_{j-1}, X_j)$. We have

$$\begin{aligned}
f_{\mathbf{X}_{1:n}}(\mathbf{x}_{1:n}) &= f_{X_1|\mathbf{X}_{2:n}}(x_1|\mathbf{x}_{2:n}) f_{X_2|\mathbf{X}_{3:n}}(x_2|\mathbf{x}_{3:n}) \cdots f_{X_{n-1}|X_n}(x_{n-1}|x_n) f_{X_n}(x_n) \\
&= f_{X_1|X_2}(x_1|x_2) \cdots f_{X_{n-1}|X_n}(x_{n-1}|x_n) f_{X_n}(x_n) \\
&= \frac{\mathbb{I}(0 < x_1 < x_2)}{x_2} \frac{\mathbb{I}(0 < x_2 < x_3)}{x_3} \cdots \frac{\mathbb{I}(0 < x_{n-1} < x_n)}{x_n} \mathbb{I}(0 < x_n < 1) \\
&= \frac{\mathbb{I}(0 < x_1 < x_2 < \cdots < x_n < 1)}{x_2 \cdots x_n}
\end{aligned}$$

Hence

$$\begin{aligned}
f_{X_1}(x_1) &= \int_{\mathbb{R}^{n-1}} \frac{\mathbb{I}(0 < x_1 < x_2 < \cdots < x_n < 1)}{x_2 \cdots x_n} d\mathbf{x}_{2:n} \\
&= \int_{x_1}^1 \int_{x_2}^1 \cdots \int_{x_{n-1}}^1 \frac{1}{x_2 \cdots x_n} dx_n \cdots dx_2 \\
&= \frac{\log^{n-1}(1/x_1)}{(n-1)!}
\end{aligned}$$

where the integral can be proven inductively. □

6.4. Additional Problems

1. Suppose $B_i \overset{\text{iid}}{\sim} \text{Ber}(p_i)$, and let $B = \sum_{i=1}^n B_i$. Unfortunately, B does **not** follow a binomial distribution. We might try to approximate B with $C \sim \text{Binom}(n, \overline{p})$, where $\overline{p} = \frac{1}{n}\sum_{i=1}^n p_i$.

   (a) Show that $M_B(t) \leq M_C(t)$ for all $t \in \mathbb{R}$. **Hint:** Use the AM-GM inequality

   $$\left(\prod_{i=1}^n a_i\right)^{\frac{1}{n}} \leq \frac{1}{n}\sum_{i=1}^n a_i$$

   for non-negative $a_1, \cdots, a_n$.

   (b) Show that $\mathbb{E}B = \mathbb{E}C$ and $\text{Var}(B) \leq \text{Var}(C)$.

2. Suppose $Y|X \sim \text{Ber}(\Phi(\alpha+\beta X))$, where $\Phi(t) = \int_{-\infty}^{t} \phi(x)dx$ is the standard normal CDF. Instead of observing X, I give you Z, which is a mis-measured version of X. Suppose $X|Z \sim N(\gamma + \delta Z, \tau^2)$.

   (a) First, prove the integral

   $$\int_{-\infty}^{\infty} \Phi(a+bx)\phi(x)dx = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right)$$

   (b) Assuming $Y|(X,Z) = Y|X$, show that $Y|Z \sim \text{Ber}(\Phi(\widetilde{\alpha} + \widetilde{\beta} Z))$, where

   $$\widetilde{\alpha} = \frac{\alpha+\beta\gamma}{\sqrt{1+\beta^2\tau^2}}, \qquad \widetilde{\beta} = \frac{\beta\delta}{\sqrt{1+\beta^2\tau^2}}$$

   In most measurement error contexts, $0 < \delta < 1$ and $\tau^2 > 0$, hence $|\widetilde{\beta}| < |\beta|$, so the mis-measured covariate provides a bias towards the null for the effect of X on Y.

3. Suppose $X_1, \cdots, X_n$ are independent RV's, and let $Z = h(X_1, \cdots, X_n)$. Let $\mathbf{X}_{-i} = (X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n)$ and $\mathbf{X}_{i:j} = (X_i, X_{i+1}, \cdots, X_{j-1}, X_j)$. Define

$$\mathbb{E}_i Z = \mathbb{E}[Z|\mathbf{X}_{1:i}] = \int_{\mathbb{R}^{n-i}} h(X_1, \cdots, X_i, x_{i+1}, \cdots, x_n) dF_{X_{i+1}}(x_{i+1}) \cdots dF_{X_n}(x_n)$$
$$\mathbb{E}^{(i)} Z = \mathbb{E}[Z|\mathbf{X}_{-i}] = \int_{\mathbb{R}} h(X_1, \cdots, X_{i-1}, x_i, X_{i+1}, \cdots, X_n) dF_{X_i}(x_i)$$

(a) Show that $\mathbb{E}_i \mathbb{E}^{(i)} Z = \mathbb{E}_{i-1} Z$.

(b) Define $\Delta_i = \mathbb{E}_i Z - \mathbb{E}_{i-1} Z$. Show that $\mathbb{E}_i \Delta_j = 0$ for $j > i$. Along with the representation $Z - \mathbb{E}Z = \sum_{i=1}^n \Delta_i$, conclude that

$$\mathrm{Var}(Z) = \sum_{i=1}^n \mathbb{E}\Delta_i^2$$

(c) Using part (a), prove that

$$\Delta_i^2 \leq \mathbb{E}_i (Z - \mathbb{E}^{(i)} Z)^2$$

(d) Use (b) and (c) to conclude that

$$\mathrm{Var}(Z) \leq \sum_{i=1}^n \mathbb{E}(Z - \mathbb{E}^{(i)} Z)^2$$

This is known as the *Efron-Stein inequality*, which provides a bound on the variance of a function of independent RV's.