Probability and Stochastic Processes
Lecture 13: Conditional Distribution

Prof. Washim Uddin Mondal
Department of Electrical Engineering
Indian Institute of Technology Kanpur

Discrete Case

Definition

If (X, Y) is a discrete random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, and $y \in \mathbb{R}$ is such that $\mathbf{P}(Y = y) > 0$, then $\forall x \in \mathbb{R}$, we have:

$$\mathbf{P}(X = x | Y = y) = \frac{\mathbf{P}(X = x, Y = y)}{\mathbf{P}(Y = y)}$$

The function $\mathbf{P}(X = \cdot | Y = y) : \mathbb{R} \to [0, 1]$ is called the conditional PMF of X given $Y = y$.

Notes

- Note that $\mathbf{P}(X = x | Y = y) \geq 0, \forall x \in \mathbb{R}$.
- $\sum_{x_i} \mathbf{P}(X = x_i | Y = y) = 1$, where $\sum_{x_i}$ is the sum over $\mathrm{range}(X)$.

Discrete Case

Example

Let $(\Omega, 2^{\Omega}, \mathbf{P})$ be a probability space such that $\Omega = \{a, b, c\}$, and $\mathbf{P}(\{a\}) = 0.5$, and $\mathbf{P}(\{b\}) = \mathbf{P}(\{c\}) = 0.25$. If $X_1, X_2 : \Omega \to \mathbb{R}$ are such that $X_1(a) = X_1(b) = 0$, $X_1(c) = 1$, and $X_2(a) = 0, X_2(b) = X_2(c) = 1$, find the conditional PMF of X_1 given $X_2 = 1$.

Solution

$$\mathbf{P}(X_1 = x \mid X_2 = 1) = \frac{\mathbf{P}(X_1 = x, X_2 = 1)}{\mathbf{P}(X_2 = 1)} = \begin{cases} \mathbf{P}(\{b\})/\mathbf{P}(\{b, c\}) \\ \mathbf{P}(\{c\})/\mathbf{P}(\{b, c\}) \\ \mathbf{P}(\{\phi\})/\mathbf{P}(\{b, c\}) \end{cases} = \begin{cases} 0.5 & \text{if } x = 0 \\ 0.5 & \text{if } x = 1 \\ 0 & \text{elsewhere} \end{cases}$$

Discrete Case

Example (Poisson Thinning)

Let the number of people entering a building be Poisson distributed with parameter λ. If each person entering the building is male with probability p, and female with probability $1-p$, independent of others, find the probability that exactly k males enter the building.

Solution

- Let X and Y respectively be the number of males and females entering the building.
- Clearly, $X + Y \sim \mathrm{Poisson}(\lambda)$.
- Note that $\mathbf{P}(X = k) = \sum_{r=0}^{\infty} \mathbf{P}(X = k, Y = r)$.

Discrete Case

Solution (Continued)

- We have the following.

$$\begin{aligned}
\mathbf{P}(X=k) &= \sum_{r=0}^{\infty} \mathbf{P}(X=k, Y=r \mid X+Y=k+r) \times \mathbf{P}(X+Y=k+r) \\
&= \sum_{r=0}^{\infty} \binom{k+r}{k} p^k (1-p)^r \times e^{-\lambda} \frac{\lambda^{k+r}}{(k+r)!} \\
&= \frac{e^{-\lambda}(\lambda p)^k}{k!} \sum_{r=0}^{\infty} \frac{[\lambda(1-p)]^r}{r!} = e^{-\lambda p} \frac{(\lambda p)^k}{k!}
\end{aligned}$$

- Hence, $X \sim \text{Poisson}(\lambda p)$.

Discrete Case

Example

Let X_1, X_2 be independent Poisson random variables with parameters λ_1, λ_2 respectively. Find the conditional PMF of X_1 given $X_1 + X_2 = n$, where $n \in \{0, 1, \cdots\}$.

Solution

- Choose $k \in \{0, 1, \cdots, n\}$. We have the following.

$$\begin{aligned}
\mathbf{P}(X_1 = k \mid X_1 + X_2 = n) &= \frac{\mathbf{P}(X_1 = k, X_1 + X_2 = n)}{\mathbf{P}(X_1 + X_2 = n)} \\
&= \frac{\mathbf{P}(X_1 = k)\mathbf{P}(X_2 = n - k)}{\mathbf{P}(X_1 + X_2 = n)}
\end{aligned}$$

Discrete Case

Solution (Continued)

- Since X_1, X_2 are independent, we have $X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$. Hence,

$$\mathbf{P}(X_1 = k \mid X_1 + X_2 = n) = \frac{e^{-\lambda_1} \dfrac{\lambda_1^k}{k!} e^{-\lambda_2} \dfrac{\lambda_2^{n-k}}{(n-k)!}}{e^{-(\lambda_1+\lambda_2)} \dfrac{(\lambda_1 + \lambda_2)^n}{n!}} = \binom{n}{k} p^k (1-p)^{n-k}$$

  where $p = \lambda_1 / (\lambda_1 + \lambda_2)$.
- The conditional distribution of X_1 given $X_1 + X_2 = n$ is Binomial (n, p).

Discrete Case

Example

Let X_1, X_2 be independent binomial random variables with parameters (m, p) and (n, p), respectively. Find the conditional PMF of X_1 given $X_1 + X_2 = r$, where $r \leq m + n$.

Solution

- Recall that $X_1 + X_2 \sim \text{Binomial}(m + n, p)$. Hence, for $k \in \{0, \cdots, r\}$, we have:

$$\begin{aligned}
\mathbf{P}(X_1 = k \mid X_1 + X_2 = r) &= \frac{\mathbf{P}(X_1 = k)\mathbf{P}(X_2 = r - k)}{\mathbf{P}(X_1 + X_2 = r)} \\
&= \frac{\binom{m}{k} p^k (1-p)^{m-k} \binom{n}{r-k} p^{r-k} (1-p)^{n-r+k}}{\binom{m+n}{r} p^r (1-p)^{m+n-r}} = \frac{\binom{m}{k}\binom{n}{r-k}}{\binom{m+n}{r}}
\end{aligned}$$

Discrete Case

Notes

- We use the notation that $\binom{m}{k} = 0$ for $k > m$ and $\binom{n}{r-k} = 0$ for $r - k > n$.
- Note that the conditional PMF is independent of p. The resulting PMF is called a Hypergeometric distribution with (m, n, r) parameters.
- Recall that $\mathrm{Bernoulli}(p)$ is the same distribution as $\mathrm{Binomial}(1, p)$. Therefore, if X_1, X_2 are independent Bernoulli random variables with parameter p, then the conditional PMF of X_1 given $X_1 + X_2 = r$ can be computed similarly for $r \in \{0, 1, 2\}$.

Discrete Case

Example

Let X_1, X_2 be independent negative binomial random variables with parameters (r, p) and (s, p), respectively. Find the conditional PMF of X_1 given $X_1 + X_2 = n$, where $n \geq r + s$.

Solution

- Recall that $X_1 + X_2 \sim \mathrm{NB}(r + s, p)$. Hence, for $k \in \{r, r + 1, \cdots\}$, we have:

$$\mathbf{P}(X_1 = k \mid X_1 + X_2 = n) = \frac{\mathbf{P}(X_1 = k)\mathbf{P}(X_2 = n - k)}{\mathbf{P}(X_1 + X_2 = n)}$$

Discrete Case

Solution (Continued)

$$\mathbf{P}(X_1 = k \mid X_1 + X_2 = r) = \frac{\binom{k-1}{r-1} p^r (1-p)^{k-r} \binom{n-k-1}{s-1} p^s (1-p)^{n-k-s}}{\binom{n-1}{r+s-1} p^{r+s} (1-p)^{n-r-s}}$$

$$= \frac{\binom{k-1}{r-1}\binom{n-k-1}{s-1}}{\binom{n-1}{r+s-1}}$$

- We use the notation that $\binom{n-k-1}{s-1} = 0$ if $n-k<s$.
- Note that the obtained PMF is independent of p.
- Recall that $\mathrm{Geom}(p)$ is the same distribution as $\mathrm{NB}(1,p)$. Hence, similar results can be obtained for independent geometric random variables.

Continuous Case

Definition

Let (X, Y) be a continuous random vector with joint PDF $f_{X,Y}$. If the marginal PDF of Y is f_Y and $y \in \mathbb{R}$ is such that $f_Y(y) > 0$, then the conditional PDF of X given $Y = y$ is defined to be a function $f_{X|Y}(\cdot|y) : \mathbb{R} \to [0, \infty)$ such that

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \ \forall x \in \mathbb{R}$$

Note

- Note that $\int_{\mathbb{R}} f_{X|Y}(x|y)\mathrm{d}x = 1$.

Continuous Case

Notes

- If the underlying probability measure is $\mathbf{P}$, then for any $B \in \mathcal{B}(\mathbb{R})$, we have:

$$\mathbf{P}(X \in B \mid Y = y) = \int_B f_{X|Y}(x|y)\mathrm{d}x$$

- Note that $\mathbf{P}(Y = y) = 0$ for a continuous random variable Y. Therefore, the above probability cannot be interpreted in the usual sense. Instead, it can interpreted as:

$$\mathbf{P}(X \in B \mid Y = y) = \lim_{\Delta y \to 0} \frac{\mathbf{P}(X \in B, y \leq Y \leq y + \Delta y)}{\mathbf{P}(y \leq Y \leq y + \Delta y)}$$

Continuous Case

Example

Let the joint PDF of (X, Y) be given as follows.

$$f_{X,Y}(x,y) = \begin{cases} y^{-1}\exp\left(-\left(\frac{x}{y}+y\right)\right) & \text{if } 0 < x < \infty, 0 < y < \infty \\ 0 & \text{elsewhere} \end{cases}$$

Find $\mathbf{P}(X > 1|Y = y)$, where $0 < y < \infty$.

Solution

- Choose $y > 0$. We have: $f_Y(y) = \int_0^\infty f_{X,Y}(x,y)\mathrm{d}x$.

Continuous Case

Solution (Continued)

$$f_Y(y) = e^{-y} \int_0^{\infty} \frac{1}{y} e^{-\frac{x}{y}} \mathrm{d}x = e^{-y}$$

- The conditional PDF of X given $Y = y$ can be expressed as follows.

$$f_{X|Y}(x|y) = \frac{1}{y} e^{-\frac{x}{y}} \mathbf{1}(x > 0)$$

- The required probability is:

$$\mathbf{P}(X > 1 | Y = y) = \int_1^{\infty} \frac{1}{y} e^{-\frac{x}{y}} \mathrm{d}x = e^{-\frac{1}{y}}$$

Continuous Case

Example (Bivariate Normal Distribution)

If the joint PDF of (X, Y) is given as follows for some $\rho \in (-1, 1)$,

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{(2\pi)^2\sigma_X^2\sigma_Y^2(1-\rho^2)}} \times \exp\left(-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-\mu_X)^2}{2\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right\}\right)$$

then prove that the conditional distribution of X given $Y = y$ is $\mathcal{N}(\mu_{X|Y}, \sigma_{X|Y}^2)$, where $\mu_{X|Y} = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y)$ and $\sigma_{X|Y}^2 = \sigma_X^2(1-\rho^2)$.

Continuous Case

Solution

- Recall that $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. Hence,

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left(-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right)$$

- Using $f_{X|Y}(x|y) = f_{X,Y}(x,y)/f_Y(y)$, we obtain the following.

$$f_{X|Y}(x|y) = \frac{1}{\sqrt{2\pi\sigma_X^2(1-\rho^2)}} \exp\left(-\frac{\left(x-\mu_X-\rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right)^2}{2(1-\rho^2)\sigma_X^2}\right)$$

Continuous Case

Example

Let X_1, X_2 be independent Gamma random variables with parameters (α_1, λ) and (α_2, λ), respectively. Find the conditional PDF of X_1 given $X_1 + X_2 = y$, where $y > 0$.

Solution

- Recall that $X_1 + X_2 \sim \mathrm{Gamma}(\alpha_1 + \alpha_2, \lambda)$.
- Note that $f_{X_1|X_1+X_2}(x_1|y) = 0$ if $x_1 \notin (0, y)$. Choose $x_1 \in (0, y)$.

$$f_{X_1|X_1+X_2}(x_1|y) = \frac{f_{X_1,X_2}(x_1, y - x_1)}{f_{X_1+X_2}(y)} = \frac{f_{X_1}(x_1) f_{X_2}(y - x_1)}{f_{X_1+X_2}(y)}$$

Continuous Case

Solution (Continued)

$$
\begin{aligned}
& f_{X_1|X_1+X_2}(x_1|y) \\
& = \frac{\lambda^{\alpha_1} x_1^{\alpha_1-1} e^{-\lambda x_1} \lambda^{\alpha_2} (y-x_1)^{\alpha_2-1} e^{-\lambda(y-x_1)}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \times \frac{\Gamma(\alpha_1+\alpha_2)}{\lambda^{\alpha_1+\alpha_2} y^{\alpha_1+\alpha_2-1} e^{-\lambda y}} \\
& = \frac{1}{B(\alpha_1,\alpha_2)} \frac{1}{y} \left(\frac{x_1}{y}\right)^{\alpha_1-1} \left(1-\frac{x_1}{y}\right)^{\alpha_2-1}
\end{aligned}
$$

- The obtained PDF is called a *scaled* Beta distribution.
- Exponential and Chi-squared are special cases of the Gamma distribution. Similar results can be obtained for these distributions using the same method.

Mixed Case

Proposition 13.1

Let X be a continuous random variable with PDF f_X, and N be a discrete random variable. If $x, n \in \mathbb{R}$ are such that $\mathbf{P}(N = n) > 0$, and $\mathbf{P}(N = n|X = x)$ is well-defined, then

$$f_{X|N}(x|n) = \frac{\mathbf{P}(N = n|X = x)f_X(x)}{\mathbf{P}(N = n)}$$

Proof of Proposition 13.1

$$\begin{aligned} &\mathbf{P}(x \leq X \leq x + \Delta x \mid N = n)\mathbf{P}(N = n) \\ &= \mathbf{P}(N = n \mid x \leq X \leq x + \Delta x)\mathbf{P}(x \leq X \leq x + \Delta) \end{aligned}$$

Mixed Case

Example

The probability of the occurrence of a head of a biased coin is modeled as a random variable P that is Beta distributed with parameters (a, b), where $a, b > 0$. If the coin is tossed $m + n$ times, and it is observed that m heads and n tails occur, what is the conditional PDF of P given this observation? Assume that the outcome of each toss is independent of that of the others.

Solution

- Let N denote the number of occurrences of head in $m + n$ trials.
- Note that $\mathbf{P}(N = m \mid P = p) = \binom{m+n}{m} p^m (1-p)^n, \forall p \in (0, 1)$.

Mixed Case

Solution (Continued)

- The PDF of P can be expressed as follows.

$$f_P(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}, \quad 0 < p < 1$$

- Moreover, $\mathbf{P}(N = m)$ can be computed as follows.

$$\begin{aligned}
\mathbf{P}(N = m) &= \int_0^1 \mathbf{P}(N = m \mid P = p) f_P(p) \mathrm{d}p \\
&= \frac{\binom{m+n}{m}}{B(a,b)} \int_0^1 p^{a+m-1}(1-p)^{b+n-1} \mathrm{d}p = \frac{\binom{m+n}{m}}{B(a,b)} B(a+m, b+n)
\end{aligned}$$

Mixed Case

Solution (Continued)

$$f_{P|N}(p|m) = \frac{\mathbf{P}(N=m|P=p)f_P(p)}{\mathbf{P}(N=m)} = \begin{cases} \dfrac{p^{a+m-1}(1-p)^{b+n-1}}{B(a+m,b+n)} & \text{if } p \in (0,1) \\ 0 & \text{elsewhere} \end{cases}$$

- The conditional distribution of P given the observation is $\mathrm{Beta}(a+m, b+n)$.
- If the prior is a beta, and the likelihood is a binomial distribution, then the posterior is again a beta distribution. Due to this reason, Beta distributions are called *conjugate priors* of the binomial likelihood.
- Applications in Thompson sampling in Bayesian RL.

Mixed Case

Example

Let $W \sim \mathrm{Gamma}(\alpha, \lambda)$. Assume that the conditional distribution of X given $W = w$, where $w > 0$, is $\mathrm{Poisson}(w)$. Show that the conditional distribution of W given $X = k$, where $k \in \{0, 1, \cdots\}$, is $\mathrm{Gamma}(\alpha + k, 1 + \lambda)$.

Solution

$$\begin{aligned}\mathbf{P}(X = k) &= \int_0^\infty \mathbf{P}(X = k | W = w) f_W(w) \mathrm{d}w \\ &= \int_0^\infty e^{-w} \frac{w^k}{k!} \frac{\lambda^\alpha w^{\alpha-1} e^{-\lambda w}}{\Gamma(\alpha)} \mathrm{d}w = \frac{\lambda^\alpha}{k!} \int_0^\infty \frac{e^{-(1+\lambda)w} w^{k+\alpha-1}}{\Gamma(\alpha)} \mathrm{d}w\end{aligned}$$

Mixed Case

Solution (Continued)

- Substituting $y = (1+\lambda)w$, we get:

$$\mathbf{P}(X=k) = \frac{\lambda^\alpha}{k!(1+\lambda)^{k+\alpha}} \int_0^\infty \frac{e^{-y} y^{k+\alpha-1}}{\Gamma(\alpha)} \mathrm{d}y = \frac{\lambda^\alpha \Gamma(k+\alpha)}{k!\Gamma(\alpha)(1+\lambda)^{k+\alpha}}$$

- Note that $f_{W|X}(w|k) = 0$ for $w \le 0$. Take $w > 0$. Using Proposition 13.1, we get:

$$f_{W|X}(w|k) = \frac{e^{-w}\frac{w^k}{k!} \frac{\lambda^\alpha w^{\alpha-1} e^{-\lambda w}}{\Gamma(\alpha)}}{\frac{\lambda^\alpha \Gamma(k+\alpha)}{k!\Gamma(\alpha)(1+\lambda)^{k+\alpha}}} = \frac{(1+\lambda)^{k+\alpha} w^{\alpha+k-1} e^{-(1+\lambda)w}}{\Gamma(k+\alpha)}$$