Conditional probability density

Recall the formula: $P[Y \text{ in } A|X = x] = \dfrac{P[Y \text{ in } A, X = x]}{P[X = x]}$.

If X is discrete, then the conditional cdf of Y given $X = x_k$ is given by

$$F_Y(y|x_k) = \frac{P[Y \le y, X = x_k]}{P[X = x_k]}, \quad P[X = x_k] > 0.$$

The conditional pdf of Y, given $X = x_k$, is given by

$$f_Y(y|x_k) = \frac{d}{dy} F_Y(y|x_k).$$

The probability of an event A given $X = x_k$ is given by

$$P[Y \text{ in } A|X = x_k] = \int_{y \text{ in } A} f_Y(y|x_k)\ dy.$$

If X and Y are independent, then

$$P[Y \le y, X = x_k] = P[Y \le y]P[X = x_k]$$

so $F_Y(y|x_k) = F_Y(y)$ and $f_Y(y|x_k) = f_Y(y)$.

X and Y are both discrete random variables

If both X and Y are discrete, then the conditional pmf of Y given $X = x_k$:

$$\begin{aligned} P_Y(y_j|x_k) &= P[Y = y_j|X = x_k] \\ &= \frac{P[X = x_k, Y = y_j]}{P[X = x_k]} = \frac{P_{XY}(x_k, y_j)}{P_X(x_k)}, \qquad P[X = x_k] > 0. \end{aligned}$$

For notational convenience, we define $P_Y(y_j|x_k) = 0$ for those x_k with $P[X = x_k] = 0$.

The probability of any event A, given $X = x_k$, is given by

$$P[Y \text{ in } A|X = x_k] = \sum_{y_j \text{ in } A} P_Y(y_j|x_k).$$

If X and Y are independent, then

$$P_Y(y_j|x_k) = \frac{P[X = x_k]P[Y = y_j]}{P[X = x_k]} = P[Y = y_j] = P_Y(y_j).$$

X and Y are both continuous random variables

If X is a continuous random variable, then $P[X = x] = 0$. Suppose X and Y are jointly continuous random variables with a joint pdf that is continuous and non-zero over some region, then the conditional cdf of Y given $X = x$ is defined by

$$F_Y(y|x) = \lim_{h \to 0} F_Y(y|x < X \leq x + h).$$

Let us consider

$$\frac{P[Y \leq y, x < X \leq x + h]}{P[x < X \leq x + h]} = \frac{\int_{-\infty}^{y} \int_{x}^{x+h} f_{XY}(x', y') \; dx' dy'}{\int_{x}^{x+h} f_X(x') \; dx'}$$

$$\approx \frac{h \int_{-\infty}^{y} f_{XY}(x, y') \; dy'}{h f_X(x)} \text{ when } h \text{ is very small.}$$

Taking $h \to 0$, $F_Y(y|x) = \dfrac{\displaystyle\int_{-\infty}^{y} f_{XY}(x,y')\ dy'}{f_X(x)}$.

The conditional pdf of Y given $X = x$ is given by

$$f_Y(y|x) = \frac{d}{dy} F_Y(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}.$$

Here, $f_Y(y|x)dy$ is the probability that Y is in the infinitesimal strip $(y, y + dy)$ given that X is in the infinitesimal strip $(x, x + dx)$. Mathematically, we have

$$\begin{aligned}
& f_Y(y|x)\ dy \\
\approx\ & P[y < Y \leq y + dy | x < X \leq x + dx] \\
=\ & \frac{P[x < X \leq x + dx, y < Y \leq y + dy]}{P[x < X \leq x + dx]} \approx \frac{f_{XY}(x,y)\ dxdy}{f_X(x)\ dx}.
\end{aligned}$$

If X and Y are independent, then $f_{XY}(x,y) = f_X(x) f_Y(y)$ so $f_Y(y|x) = f_Y(y)$ and $F_Y(y|x) = F_Y(y)$.

Example Suppose the joint pdf of X and Y is given by

$$f_{XY}(x,y) = \begin{cases} 2e^{-x}e^{-y} & 0 \leq y \leq x < \infty \\ 0 & \text{otherwise} \end{cases},$$

find $f_X(x|y)$ and $f_Y(y|x)$.

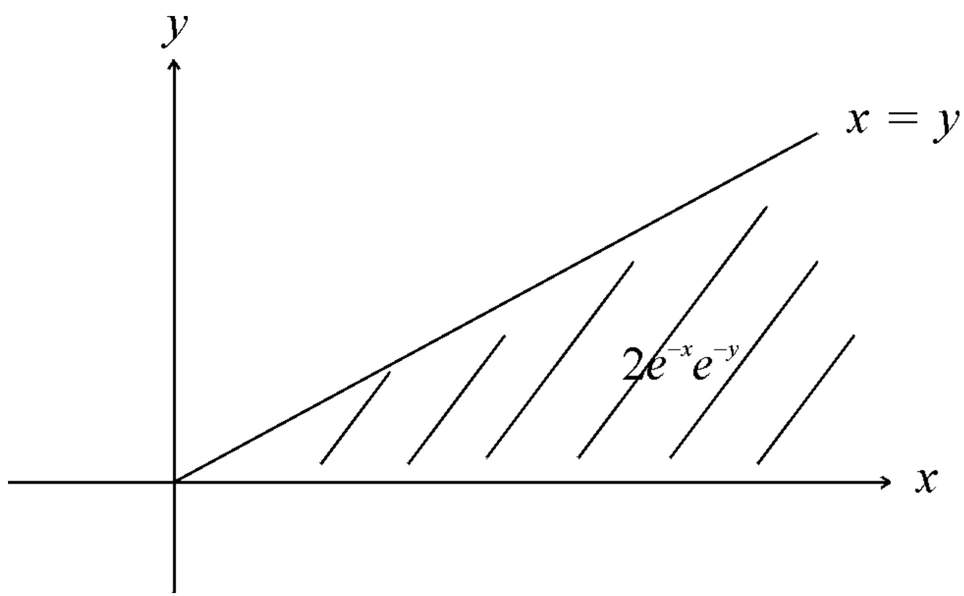
Solution

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x,y')\ dy' \\ &= \int_0^x 2e^{-x}e^{-y'}dy' = 2e^{-x}(1-e^{-x}), \qquad 0 \leq x < \infty, \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x',y)\ dx' = \int_y^{\infty} 2e^{-x'}e^{-y}\ dx' = 2e^{-2y}, \qquad 0 \leq y < \infty, \end{aligned}$$

then

$$f_X(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{2e^{-x}e^{-y}}{2e^{-2y}} = e^{-(x-y)} \quad \text{for } 0 \le y \le x,$$

$$f_Y(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{2e^{-x}e^{-y}}{2e^{-x}(1-e^{-x})} = \frac{e^{-y}}{1-e^{-x}} \quad \text{for } 0 \le y \le x.$$



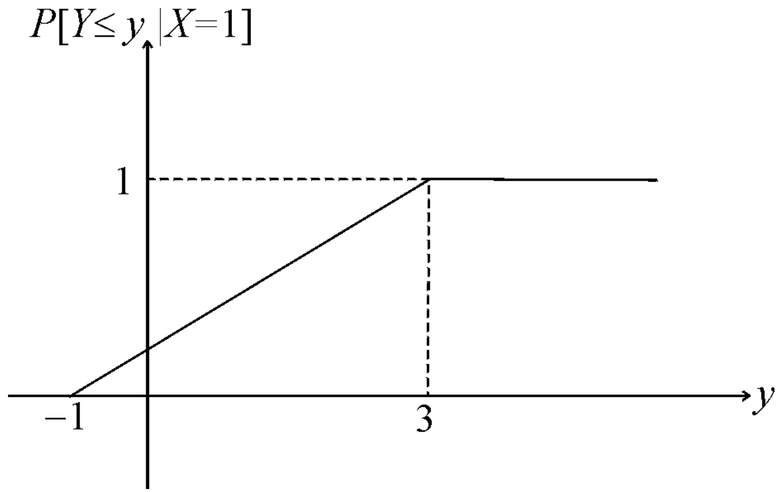
Example

X = input voltage, Y = output = input + noise voltage.

Here, $X = 1$ or -1 with equal probabilities, the noise is uniformly distributed from -2 to 2 with equal probabilities.

When $X = 1$, Y becomes uniformly distributed in $[-1, 3]$ so

$$f_Y(y|X=1) = \begin{cases} \frac{1}{4} & -1 \leq y \leq 3 \\ 0 & \text{otherwise} \end{cases}.$$



The conditional cdf of Y:

$$\begin{aligned}
F_Y(y|X=1) &= P[Y \le y|X=1] = \int_{-\infty}^{y} f_Y(y'|X=1)\ dy' \\
&= \begin{cases} 0, & -1 > y \\ \displaystyle\int_{-1}^{y} \frac{1}{4}\ dy' = \frac{y+1}{4}, & -1 \le y \le 3 \\ 1, & y > 3 \end{cases} .
\end{aligned}$$

For example,

$$\begin{aligned}
P[Y \le 0|X=1] &= \int_{-1}^{0} f_Y(y'|X=1)\ dy' = F_Y(0|X=1) = \frac{1}{4}, \\
P[1 \le Y \le 2|X=1] &= F_Y(2|X=1) - F_Y(1|X=1) = \frac{3}{4} - \frac{2}{4} = \frac{1}{4}, \\
P[Y \le 0, X=1] &= P[Y \le 0|X=1]P[X=1], \\
&= \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}.
\end{aligned}$$

It can be seen easily that

$$F_Y(y|X=-1) = \begin{cases} 0 & -3 > y \\ \frac{y+3}{4} & -3 \le y \le 1 \\ 1 & y > 1 \end{cases} .$$

Example

A point (X, Y) is selected at random inside the unit circle.

(a) Find the marginal pdf of $X, f_X(x)$.

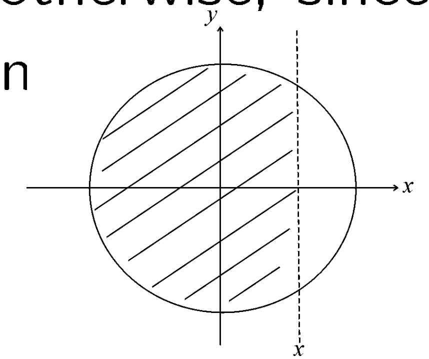
(b) Find the conditional pdf of X given $Y, f_X(x|y)$.

(c) Are X and Y independent?

Solution

(a) Marginal cdf of $X = F_X(x) = P[X \le x] = \int_{-1}^{x} \int_{-\sqrt{1-x'^2}}^{\sqrt{1-x'^2}} f(x', y')\ dy'dx'$.

Note that $f(x,y) = \dfrac{1}{\pi}$ for $x^2 + y^2 \le 1$, and $f(x,y) = 0$ if otherwise; since (X,Y) is selected at random from the unit circle, we obtain



$$F_X(x) = \frac{1}{\pi} \int_{-1}^{x} 2\sqrt{1 - x'^2}\ dx'.$$

Marginal pdf of $X = f_X(x) = \frac{d}{dx} F_X(x)$

$$= \frac{d}{dx} \int_{-1}^{x} \frac{2\sqrt{1 - x'^2}}{\pi}\ dx' = \frac{2}{\pi}\sqrt{1 - x^2}.$$

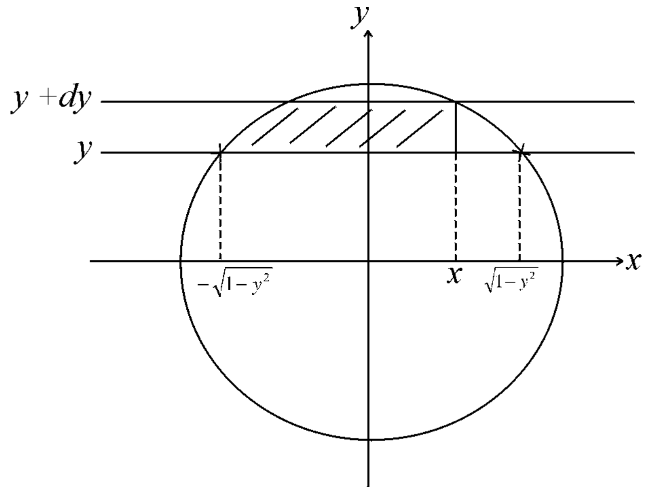
Indeed, $f_X(x)\ dx = P[x < X \le x + dx] = \dfrac{\text{area of the strip}}{\text{area of circle}} = \dfrac{2\sqrt{1 - x^2}\ dx}{\pi}$.

(b) The conditional cdf of X is given by

$$F_X(x|y) = \lim_{h \to 0} \frac{P[X \leq x, y < Y \leq y + h]}{P[y < Y \leq y + h]}$$

$$= \lim_{h \to 0} \frac{h[x - (-\sqrt{1 - y^2})]}{h[\sqrt{1 - y^2} - (-\sqrt{1 - y^2})]},$$

$$= \frac{x + \sqrt{1 - y^2}}{2\sqrt{1 - y^2}}$$



so that

$$f_X(x|y) = \frac{d}{dx}\left[\frac{x + \sqrt{1 - y^2}}{2\sqrt{1 - y^2}}\right] = \frac{1}{2\sqrt{1 - y^2}}.$$

(c) X and Y are dependent since $f_X(x|y) \neq f_X(x)$. Is it obvious from intuition?

Remark

1. As a verification, we may use the relation

$$f_{XY}(x,y) = f_X(x|y)f_Y(y)$$

to compute $f_X(x|y)$. Note that $f_{XY}(x,y) = \frac{1}{\pi}$ and $f_Y(y) = \frac{2}{\pi}\sqrt{1-y^2}$.

Hence, $f_X(x|y) = \dfrac{\frac{1}{\pi}}{\frac{2}{\pi}\sqrt{1-y^2}} = \dfrac{1}{2\sqrt{1-y^2}}$.

2.
$$\begin{aligned}
f_X(x|y)\ dx &= P[x < X \le x + dx | y < Y \le y + dy] \\
&= \frac{\text{area of differential patch, } dxdy}{\text{area of horizontal strip within } (y, y+dy)} \\
&= \frac{dxdy}{2\sqrt{1-y^2}\ dy} = \frac{1}{2\sqrt{1-y^2}}\ dx.
\end{aligned}$$

Example

If X and Y are independent Poisson random variables with respective parameters λ_1 and λ_2. Show that $X+Y$ has a Poisson distribution with parameter $\lambda_1+\lambda_2$. Then calculate the conditional pmf of X, given that $X+Y=n$.

Solution

$$\begin{aligned}
P[X+Y=n] &= \sum_{k=0}^{n} P[X=k, Y=n-k] = \sum_{k=0}^{n} P[X=k]P[Y=n-k] \\
&= \sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n.
\end{aligned}$$

Next, the conditional pmf of X, given $X+Y=n$, is given by

$$\begin{aligned} P[X=k|X+Y=n] &= \frac{P[X=k, X+Y=n]}{P[X+Y=n]} \\ &= \frac{P[X=k, Y=n-k]}{P[X+Y=n]} \\ &= \frac{P[X=k]P[Y=n-k]}{P[X+Y=n]}. \end{aligned}$$

Combining the relations together,

$$\begin{aligned} P[X=k|X+Y=n] &= \frac{e^{-\lambda_1}\lambda_1^k}{k!}\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\left[\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}\right]^{-1} \\ &= \frac{n!}{(n-k)!k!}\frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n} \\ &= {}_nC_k\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}. \end{aligned}$$

In other words, the conditional distribution of X, given that $X+Y=n$, is the binomial distribution with parameters n and $\lambda_1/(\lambda_1+\lambda_2)$.

Restatement of the theorem on total probability

Recall: $P[X = x_k, Y = y_j] = P[Y = y_j | X = x_k] P[X = x_k]$

or $P_{XY}(x_k, y_j) = P_Y(y_j | x_k) P_X(x_k)$.

How to find the probability that Y is in A?

$$\begin{aligned} P[Y \text{ in } A] &= \sum_{\text{all } (x_k, y_j)} \sum_{\text{in } A} P_{XY}(x_k, y_j) \\ &= \sum_{\text{all } (x_k, y_j)} \sum_{\text{in } A} P_Y(y_j | x_k) P_X(x_k) \\ &= \sum_{\text{all } x_k} P_X(x_k) \sum_{y_j \text{ in } A} P_Y(y_j | x_k), \end{aligned}$$

so $P[Y \text{ in } A] = \sum_{\text{all } x_k} P[Y \text{ in } A | X = x_k] P_X(x_k)$.

Interpretation

To compute $P[Y \text{ in } A]$, we first compute $P[Y \text{ in } A|X = x_k]$ and then average over x_k.

The above formula holds even when X is discrete, Y is continuous.

What happens when both X and Y are continuous?

$$f_{XY}(x, y) = f_Y(y|x) f_X(x)$$

$$P[Y \text{ in } A] = \int_{-\infty}^{\infty} P[Y \text{ in } A|X = x] f_X(x) \, dx.$$

The formula remains valid even when Y is discrete.

Example The total number of defects X on a chip is a Poisson random variable with mean α. Suppose that each defect has a probability p of falling in a specific region R and that the location of each defect is independent of the locations of all other defects. Find the pmf of the number of defects Y that fall in the region R.

Solution Recall the formula: $P[Y = j] = \sum_{k=0}^{\infty} P[Y = j|X = k]P[X = k]$.

Imagine performing a Bernoulli trial each time a defect occurs with a "success" occurring when the defect falls in the region R. If the total number of defects is $X = k$, then

$$P[Y = j|X = k] = \begin{cases} 0 & j > k \\ {}_kC_j p^j (1-p)^{k-j} & 0 \leq j \leq k \end{cases}.$$

Note that the summation over k is performed with $k \geq j$, so we have

$$
\begin{aligned}
P[Y=j] &= \sum_{k=j}^{\infty} \frac{k!}{j!(k-j)!} p^j (1-p)^{k-j} \frac{\alpha^k}{k!} e^{-\alpha} \\
&= \frac{(\alpha p)^j e^{-\alpha}}{j!} \sum_{k=j}^{\infty} \frac{[(1-p)\alpha]^{k-j}}{(k-j)!} \\
&= \frac{(\alpha p)^j e^{-\alpha}}{j!} e^{(1-p)\alpha} = \frac{(\alpha p)^j}{j!} e^{-\alpha p}.
\end{aligned}
$$

Hence, Y is a Poisson random variable with mean αp.

Example – Number of arrivals during a customer's service time

The number of customers that arrive at a service station within $[0, t]$ is a Poisson random variable with parameter βt. The time required to service each customer is an exponential random variable with parameter α. Find the pmf for the number of customers N that arrive during the service time T of a specific customer, assuming that customer arrivals are independent of the customer service time.

Solution

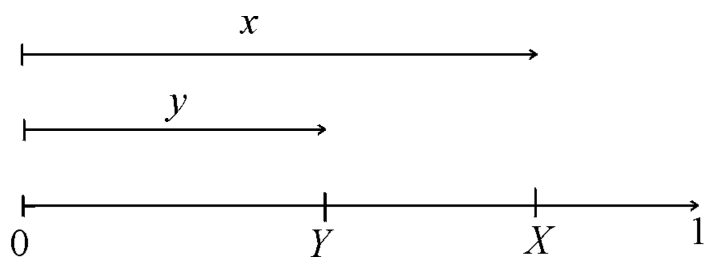
Here, N is a discrete random variable and T is a continuous random variable.

$$\begin{aligned} P[N=k] &= \int_0^\infty P[N=k|T=t] f_T(t)\ dt \\ &= \int_0^\infty \frac{(\beta t)^k}{k!} e^{-\beta t} \alpha e^{-\alpha t}\ dt \\ &= \frac{\alpha \beta^k}{k!} \int_0^\infty t^k e^{-(\alpha+\beta)t}\ dt \\ &= \frac{\alpha \beta^k}{k!(\alpha+\beta)^{k+1}} \int_0^\infty r^k e^{-r} dr, \qquad r = (\alpha+\beta)t \\ &= \frac{\alpha \beta^k}{(\alpha+\beta)^{k+1}} = \frac{\alpha}{\alpha+\beta} \left(\frac{\beta}{\alpha+\beta} \right)^k . \end{aligned}$$

Hence, N is a geometric random variable with the probability of "success" $= \dfrac{\beta}{\alpha+\beta}$.

Example The random variable X is selected at random from the unit interval. The random variable Y is then selected at random from $(0, X)$. Find the cdf and pdf of Y.

Solution



$$F_Y(y) = P[Y \le y] = \int_0^1 P[Y \le y | X = x] f_X(x)\ dx.$$

Note that $P[Y \le y | X = x] = \begin{cases} y/x & 0 < y \le x \le 1 \\ 1 & 0 \le x < y \end{cases}$,

so $F_Y(y) = \int_0^y dx' + \int_y^1 \frac{y}{x'}\ dx' = y - y \ln y$

and $f_Y(y) = -\ln y,\ \ 0 < y \le 1.$

Conditional Expectation

The conditional expectation of Y given $X = x$ is given by

$$E[Y|x] = \int_{-\infty}^{\infty} y f_Y(y|x)\ dy.$$

When X and Y are both discrete random variables

$$E[Y|x] = \sum_{y_j} y_j P_Y(y_j|x).$$

On the other hand, $E[Y|x]$ can be viewed as a function of x:

$$g(x) = E[Y|x].$$

Correspondingly, this gives rise to the random variable: $g(X) = E[Y|X]$.

What is $E[E[Y|X]]$?

Note that $E[E[Y|X]] = \begin{cases} \displaystyle\int_{-\infty}^{\infty} E[Y|x]f_X(x)\ dx, & X \text{ is continuous} \\ \displaystyle\sum_{x_k} E[Y|x_k]P_X(x_k), & X \text{ is discrete} \end{cases}.$

Suppose X and Y are jointly continuous random variables

$$\begin{aligned} E[E[Y|X]] &= \int_{-\infty}^{\infty} E[Y|x]f_X(x)\ dx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf_Y(y|x)\ dy\ f_X(x)\ dx \\ &= \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{XY}(x,y)\ dxdy \\ &= \int_{-\infty}^{\infty} yf_Y(y)\ dy = E[Y]. \end{aligned}$$

Generalization $E[h(Y)] = E[E(h(Y)|X]]$ [in the above proof, change y to $h(y)$];

and in particular, $E[Y^k] = E[E[Y^k|X]]$.

Example

A customer entering a service station is served by serviceman i with probability $p_i, i = 1, 2, \cdots, n$. The time taken by serviceman i to service a customer is an exponentially distributed random variable with parameter α_i. Let I be the discrete random variable which assumes the value i if the customer is serviced by the ith serviceman, and let $P_I(i)$ denote the probability mass function of I. Let T denote the time taken to service a customer.

(a) Explain the meaning of the following formula

$$P[T \leq t] = \sum_{i=1}^{n} P_I(i) P[T \leq t | I = i].$$

Use it to find the probability density function of T.

(b) Use the conditional expectation formula

$$E[E[T|I]] = E[T]$$

to compute $E[T]$.

Solution

(a) From the conditional probability formula, we have

$$P[T \leq t, I = i] = P_I(i)P[T \leq t | I = i].$$

The marginal distribution function $P[T \leq t]$ is obtained by summing the joint probability values $P[T \leq t, I = i]$ for all possible values of i. Hence,

$$P[T \leq t] = \sum_{i=1}^{n} P_I(i)P[T \leq t | I = i].$$

Here, $P_I(i) = p_i$ and $P[T \leq t | I = i] = 1 - e^{-\alpha_i t}, t \geq 0$. The probability density function of T is given by

$$f_T(t) = \frac{d}{dt}P[T \leq t] = \begin{cases} \sum_{i=1}^{n} p_i \alpha_i e^{-\alpha_i t} & t \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

(b)

$$\begin{aligned} E[T] &= E[E[T|I]] = \sum_{i=1}^{n} P_I(i)E[T|I=i] \\ &= \sum_{i=1}^{n} p_i \int_0^{\infty} \alpha_i t e^{-\alpha_i t}\, dt \\ &= \sum_{i=1}^{n} \frac{p_i}{\alpha_i}. \end{aligned}$$

The mean service time is the weighted average of mean service times at different counters, where $\frac{1}{\alpha_i}$ is the mean service time for the ith serviceman.