1 Reading

4.1. Derived Distributions

Given the PDF of X, we are interested in calculating the PDF of $Y = g(X)$. To do so, we should first calculate the CDF F_Y of Y using the formula

$$F_Y(y) = P(g(X) \leq y) = \int_{\{x|g(x)\leq y\}} f_X(x)dx$$

Then we can differentiate to obtain the PDF of Y:

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

As a special case, if $Y = aX + b$ then we can calculate its PDF as

$$f_Y(y) = \frac{1}{|a|}f_X\left(\frac{y-b}{a}\right)$$

4.1.1. Sum of Independent Random Variables – Convolution

We consider the random variable $Z = X + Y$ (for independent X and Y). To start, we will derive the PMF for the discrete case.

$$\begin{aligned} p_Z(z) &= P(X + Y = z) \\ &= \sum_x P(X = x, Y = z - x) \\ &= \sum_x p_X(x)p_Y(z - x) \end{aligned}$$

The PMF p_Z is called the **convolution** of the PMFs of X and Y.

For the continuous case, we can first find the joint PDF of X and Z, then integrate to find the PDF of Z. We have

$$f_{X,Z}(x, z) = f_X(x)f_{Z|X}(z|x) = f_X(x)f_Y(z - x)$$

from which we can obtain

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Z}(x, z)dx = \int_{-\infty}^{\infty} f_X(x)f_Y(z - x)dx$$

4.2. Covariance and Correlation

The **covariance** of two random variables X and Y, denoted $cov(X, Y)$, is defined as

$$cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

When $cov(X, Y) = 0$, we can say that X and Y are **uncorrelated**. Roughly speaking, a positive covariance indicates that the values of $X - \mathbb{E}[X]$ and $Y - \mathbb{E}[Y]$ obtained in a single experiment “tend” to have the same sign.

For any random variables X, Y, and Z, and any scalars a and b, we have

- $cov(X, X) = var(X)$
- $cov(X, aY + b) = a \cdot cov(X, Y)$
- $cov(X, Y + Z) = cov(X, Y) + cov(X, Z)$

Note that if X and Y are independent, we have $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, which implies that $cov(X, Y) = 0$. Therefore, if X and Y are independent then they are also uncorrelated. The converse is not necessarily true.

The **correlation coefficient** $\rho(X, Y)$ of two random variables X and Y that have nonzero variances is defined as

$$\rho(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}}$$

This is a normalized version of the covariance, and ranges from -1 to 1. (If $\rho > 0$, then the values of $X - \mathbb{E}[X]$ and $Y - \mathbb{E}[Y]$ "tend" to have the same sign, and the size of $|\rho|$ provides a normalized measure of the extent to which this is true.)

4.2.1. Variance of the Sum of Random Variables

If $X_1, ..., X_n$ are random variables with finite variance, then

$$var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} var(X_i) + \sum_{\{(i,j)\ |\ i \neq j\}} cov(X_i, X_j)$$

In the case of two random variables, we have

$$var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$$

2 Lecture

Recap: Exponential Distribution

The best way to remember the exponential is through the complementary CDF $P(X > x)$, which equals $e^{-\lambda x}$.

The exponential distribution exhibits the **memoryless property**. The memoryless property for a continuous random variable suggests that if we've waited t seconds, the probability that we have to wait s more is as good as starting from scratch.

Derived Distributions

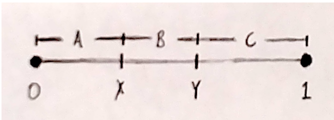
Let $X, Y \sim \text{Unif}[0, 1]$ be two independent random variables. We want to characterize $Z = \min(X, Y)$. What is its PDF $f_Z(z)$, and what is its mean $\mathbb{E}[Z]$? We have

$$\begin{aligned} P(Z > u) &= P(X > u, Y > u) \\ &= P(X > u)P(Y > u) \\ &= (1 - u)^2 \end{aligned}$$

Thus we can find the CDF by complementing it, and then just take the derivative to get the density. $F_Z(u) = 1 - (1 - u)^2$, and then $f_z(u) = \frac{d}{dz}[F_Z(u)] = 2(1 - u)$ if $0 < u < 1$ (and 0 otherwise).

From this, we have $\mathbb{E}[Z] = \int_0^1 u f_Z(u) du = \int_0^1 2u(1 - u) du = ... = 1/3$.

What is an alternate way of computing the mean? Draw a picture!



We know that $A + B + C = 1$, so let's do linearity of expectation. $\mathbb{E}[A] + \mathbb{E}[B] + \mathbb{E}[C] = 1$, and by symmetry $\mathbb{E}[A] = \mathbb{E}[B] = \mathbb{E}[C]$, so $\mathbb{E}[A] = \frac{1}{3}$. (There is no reason that any interval should be bigger than the others.)

Example. Assume that $X, Y \sim \text{Exp}(1)$ are independent. Then let $Z = \max(X,Y) - \min(X,Y)$. How is Z distributed (i.e. $Z \sim$ what)?

Suppose $X < Y$. Then, by the memoryless property,

$$\begin{aligned} P(Z > z \mid X = l) &= P(Y > z + l \mid X = l) \\ &= P(Y > z + l \mid Y > l) \\ &= P(Y > z) \end{aligned}$$

In other words, since the left-hand side depends on l but the right-hand side does not, Z does not care about l! Hence Z and X are independent. $P(Z > z) = P(Y > z) = e^{-z}$. (If $Y < X$, we get the same result.)

Thus we have $Z \sim \text{Exp}(1)$.

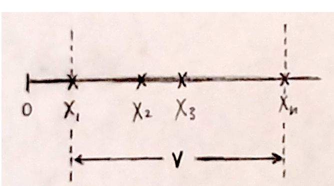
Example. Let $M = \min(X,Y)$, where $X \sim \text{Exp}(\lambda_1)$ and $Y \sim \text{Exp}(\lambda_2)$ are independent. $M \sim$ what?

We have

$$\begin{aligned} P(M > u) &= P(X > u, Y > u) \\ &= P(X > u)P(Y > u) \\ &= e^{-\lambda_1 u} e^{-\lambda_2 u} \\ &= e^{-(\lambda_1 + \lambda_2)u} \end{aligned}$$

which implies

$$M \sim \text{Exp}(\lambda_1 + \lambda_2)$$



Can we write $\max\{X_1, ..., X_n\}$ in terms of $\min\{X_1, ..., X_n\}$? Let the X_i's be i.i.d. and drawn from Exp(1).

Define $A_n = \mathbb{E}[\max\{X_1, ..., X_n\}]$. Then

$$A_n = \mathbb{E}[\min\{X_1, ..., X_n\}] + \mathbb{E}[V]$$

where V is the max of $(n-1)$ i.i.d. Exp(1) random variables. The memoryless property allows us to say this, and to build the recursion. Continuing,

$$\mathbb{E}[\min\{X_1, ..., X_n\}] + \mathbb{E}[V] = \frac{1}{n} + A_{n-1}$$

Hence $A_n = \frac{1}{n} + A_{n-1}$.

Normal Distribution

This is the most celebrated of all distributions, the mother of all distributions. The PDF of a random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

The standard normal $Z \sim \mathcal{N}(0, 1)$ has a PDF of

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Normality is preserved by linear transformations (so linear systems are nice!). Formally, if $Y = aX + b$ and $X \sim \mathcal{N}(\cdot)$, then $Y \sim \mathcal{N}(\cdot)$.

Given $F_X(x), f_X(x)$, and a linear function $Y = aX + b$, what are $F_Y(y)$ and $f_Y(y)$? We impose the constraint $a > 0$.

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(aX + b \leq y) \\
&= P\left(X \leq \frac{y-b}{a}\right) \\
&= F_X\left(\frac{y-b}{a}\right)
\end{aligned}$$

Thus when $a > 0$, $F_Y(y) = F_X\left(\frac{y-b}{a}\right)$. To find $f_Y(y)$, we take the derivative and find that $f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$.

When $a < 0$, $f_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right)$. Thus, in total, $f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$.

Suppose $X \sim \mathcal{N}(0, 1)$, and $Y = \mu + \sigma X$. Then $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ and $f_Y(y) = \frac{1}{\sigma} f_X\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2}$. In other words, $Y \sim \mathcal{N}(\mu, \sigma^2)$.

Note that $\mathbb{E}[\mathcal{N}(\mu, \sigma^2)] = \mu + \sigma$. Using calculus, we can show that $\mathbb{E}[X] = 0$ and $var(X) = 1$. Since $Y = \mu + \sigma X$, $\mathbb{E}[Y] = \mu + \sigma\mathbb{E}[X] = \mu$ and $var(Y) = \sigma^2 var(X) = \sigma^2$.

We define $\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx$. In other words, $\Phi(z)$ is the complementary CDF of the standard normal. We have $\Phi(-z) = 1 - \Phi(z)$.

Convolution

If $Z = X + Y$, where X and Y are independent,

- **Discrete setting:** Given $\{P(X = k)\}$ and $\{P(Y = k)\}$, we would like to find $\{P(Z = k)\}$.
- **Continuous setting:** Given $f_X(x)$ and $f_Y(y)$, we would like to find $f_Z(z)$.

In the discrete case,

$$\begin{aligned}
P(Z = z) &= P(X + Y = z) \\
&= \sum_{\{x,y \ | \ x+y=z\}} P(X = x, Y = y) \\
&= \sum_{x} P(X = x) P(Y = z - x)
\end{aligned}$$

In the continuous case,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx$$

The key idea is that if we add two random variables, the randomness of the sum gets larger; we are spreading out the uncertainty. **Convolution** captures how *much* we spread out the uncertainty.

References

[1] D.P. Bertsekas and J.N. Tsitsiklis. *Introduction to Probability.* Athena Scientific books. Athena Scientific, 2002.