Properties of random variables and limit theorems

1 Iterated expectations

a. Given $\mathrm{E}[X|Y] = 2Y$ and $f(Y) = .5$ with $Y \in [-3, -1]$, what is $\mathrm{E}[X]$?

Solution:

$$
\begin{aligned}
\mathrm{E}[X] &= \mathrm{E}_Y[\mathrm{E}_{Y|X}[X|Y]] \\
&= \int_{-3}^{-1} \mathrm{E}[X|Y] \Pr(Y) dy \\
&= \int_{-3}^{-1} 2 \times Y(.5) dy \\
&= \frac{Y^2}{2}|_{-3}^{-1} \\
&= \frac{1}{2} - \frac{9}{2} \\
&= -\frac{8}{2} \\
&= -4
\end{aligned}
$$

b. Given $\mathrm{E}[Z|H] = 15H - 10$ and $H \sim \text{Bernoulli}(.2)$, what is $\mathrm{E}[Z]$?

Solution:

$$
\begin{aligned}
\mathrm{E}[Z] &= \mathrm{E}_H[\mathrm{E}_{Z|H}[Z|H]] \\
&= \sum_h \mathrm{E}[Z|H = h]\ p(h) \\
&= \mathrm{E}[Z|H = 1] \Pr(H = 1) + \mathrm{E}[Z|H = 0] \Pr(H = 0) \\
&= 5 \times .2 + -10 \times .8 \\
&= 1 - 8 \\
&= -7
\end{aligned}
$$

2 Moment generating functions

Consider a variable $Y = \log(X)$ that is distributed $N(\mu, \sigma^2)$. Derive $\mathrm{E}[X]$ and $\mathrm{Var}[X]$. (**Hint**: the moment generating function of a normal random variable is $\exp(\mu t + 0.5\sigma^2 t^2)$.)

Solution:

The m.g.f. of Y is $\mathrm{E}[e^{tY}] = \mathrm{E}[e^{t\log(X)}] = \mathrm{E}[e^{\log(X)^t}] = \mathrm{E}[X^t] = \exp(\mu t + 0.5\sigma^2 t^2)$. This implies that we can find the t^{th} moment of X using the above expression. So:

$$
\begin{aligned}
\mathrm{E}[X] &= \exp(\mu \times 1 + 0.5\sigma^2 \times 1^2) = \exp(\mu + 0.5\sigma^2) \\
\mathrm{Var}(X) &= \mathrm{E}[X^2] - \mathrm{E}[X]^2 \\
&= \exp(\mu \times 2 + 0.5\sigma^2 \times 2^2) - (\exp(\mu + 0.5\sigma^2))^2 = \exp(2\mu + 2\sigma^2) - \exp(2\mu + \sigma^2) \\
&= \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]
\end{aligned}
$$

3 Changing coordinates

Suppose $f(X) = 3x^2$ with $X \in [0,1]$ and $Y = X^2$.

a. What is $f(Y)$?
b. What is $\mathrm{E}[Y]$?

Solutions:

$$
\begin{aligned}
f(X) &= 3x^2 \text{ for } X \in [0,1] \\
Y &= X^2 \text{ for } X \ \in \ [0,1] \\
X &= \sqrt{Y} \text{ for } Y \in [0,1] \\
g^{-1}(Y) &= \sqrt{Y} \\
\left|\frac{\partial}{\partial Y} g^{-1}(Y)\right| &= \frac{1}{2\sqrt{Y}} \\
f_Y(Y) &= f_x[g^{-1}(Y)]|\frac{\partial}{\partial y} g^{-1}(Y)| \\
&= 3(g^{-1}(Y))^2|\frac{1}{2\sqrt{Y}}| \\
&= 3(\sqrt{Y})^2 \frac{1}{2\sqrt{Y}} \\
&= \frac{3Y}{2\sqrt{Y}} \\
&= \frac{3}{2}\sqrt{Y} \text{ for } y \in [0,1]
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}[Y] &= \int_0^1 y f(y) dy \\
&= \int_0^1 y \frac{3}{2}\sqrt{y}\ dy \\
&= \int_0^1 \frac{3}{2} y^{3/2}\ dy \\
&= \frac{3}{5} y^{5/2}|_0^1 \\
&= \frac{3}{5}
\end{aligned}
$$

4 Markov and Chebyshev inequalities

A statistician wants to estimate the mean height h (in meters) fo a population, based on n independent samples $X_1, \ldots, X_n$, chosen uniformly from the entire population. She uses the sample mean $M_n = \frac{X_1 + \ldots + X_n}{n}$ as the estimate of h, and a rough guess of 1.0 meters for the standard deviation of the samples X_i.

a. How large should n be so that the standard deviation of M_n is at most 1 centimeter?

Solution:

We have $\sigma_{M_n} = \frac{1}{\sqrt{n}}$, so in order that $\sigma_{M_n} \leq 0.01$, we must have $n \geq 10,000$.

b. How large should n be so that Chebyshev's inequality guarantees that the estimate is within 5 centimeters from h, with probability at least 0.99?

Solution:

We want to have

$$\Pr(|M_n - h| \leq 0.05) \geq 0.99$$

Using the knowledge that $h = \mathrm{E}[M_n]$, $\sigma^2_{M_n} = \frac{1}{n}$, and the Chebyshev inequality, we have

$$\begin{aligned}
\Pr(|M_n - h| \leq 0.05) &= \Pr(|M_n - \mathrm{E}[M_n]| \leq 0.05) \\
&= 1 - \Pr(|M_n - \mathrm{E}[M_n]| \geq 0.05) \\
&\geq 1 - \frac{1/n}{(0.05)^2}
\end{aligned}$$

Thus we must have

$$1 - \frac{1/n}{(0.05)^2} \geq 0.99$$

which yields $n \geq 40,000$.

5 Gambling in the casino

Before starting to play roulette in a casino, you want to look for biases that you can exploit. You therefore watch 100 rounds that result in a number between 1 and 36, and count the number of rounds for which the result is odd. If the count exceeds 55, you decide that the roulette is not fair. Assuming that the roulette is fair, find an approximation for the probability that you will make the wrong decision.

Hint: Use a standard normal distribution CDF table.

Solution:

Let S be the number of times that the result was odd, which is a binomial random variable with parameters $n = 100$ and $p = 0.5$, so that $\mathrm{E}[X] = 100 \times 0.5 = 50$ and $\sigma_S = \sqrt{100 \times 0.5 \times 0.5} = \sqrt{25} = 5$. Using the normal approximation to the binomial, we find

$$\begin{aligned}
\Pr(S > 55) &= \Pr(\frac{S - 50}{5} > \frac{55 - 50}{5}) \\
&\approx 1 \ \Phi(1) \\
&= 1 - 0.8413 \\
&= 0.1587
\end{aligned}$$

6 Making some widgets

A factory produces X_n widgets on day n, where X_n are independent and identically distributed random variables, with mean 5 and variance 9.

a. Find an approximation to the probability that the total number of gadgets produced in 100 days is less than 440.

Solution:

Let $S_n = X_1 + \ldots + X_n$ be the total number of widgets produced in n days. Note that the mean, variance, and standard deviation of S_n is $5n$, $9n$, and $3\sqrt{n}$ respectively. Thus,

$$\begin{aligned}
\Pr(S_{100} < 440) &= \Pr\left(\frac{S_{100} - 500}{30} < \frac{439.5 - 500}{30}\right) \\
&\approx \Phi\left(\frac{439.5 - 500}{30}\right) \\
&= \Phi(-2.02) \\
&= 1 - \Phi(2.02) \\
&= 1 - 0.9783 \\
&= 0.0217
\end{aligned}$$

b. Find (approximately) the largest value of n such that

$$\Pr(X_1 + \ldots + X_n \geq 200 + 5n) \leq 0.05$$

Solution:

The requirement $\Pr(X_1 + \ldots + X_n \geq 200 + 5n) \leq 0.05$ translates to

$$\Pr\left(\frac{S_n - 5n}{3\sqrt{n}} \geq \frac{200}{3\sqrt{n}}\right) \leq 0.05$$

or, using a normal approximation,

$$\begin{aligned}
1 - \Phi\left(\frac{200}{3\sqrt{n}}\right) &\leq 0.05 \\
\Phi\left(\frac{200}{3\sqrt{n}}\right) &\geq 0.95
\end{aligned}$$

From the normal tables, we obtain $\Phi(1.65) \approx 0.95$, and therfore

$$\frac{200}{3\sqrt{n}} \geq 1.65$$

which finally yields $n \leq 1632$.

c. Let N be the first day on which the total number of widgets produced exceeds 1000. Calculate an approximation to the probability that $N \geq 220$.

Solution:

The event $N \geq 220$ (it takes at least 220 days to exceed 1000 widgets) is the same as the event $S_{219} \leq 1000$ (no more than 1000 widgets produced in the first 219 days). Thus,

$$
\begin{aligned}
\Pr(N \geq 220) &= \Pr(S_{219} \leq 1000) \\
&= \Pr\left(\frac{S_{219} - 5 \times 219}{3\sqrt{219}} \leq \frac{1000 - 5 \times 219}{3\sqrt{219}}\right) \\
&= 1 - \Phi(2.14) \\
&= 1 - 0.9838 \\
&= 0.0162
\end{aligned}
$$