

Financial Engineering & Risk Management

Review of Basic Probability

M. Haugh G. Iyengar

Department of Industrial Engineering and Operations Research
Columbia University

Discrete Random Variables

Definition. The **cumulative distribution function** (CDF), $F(\cdot)$, of a random variable, X , is defined by

$$F(x) := P(X \leq x).$$

Definition. A discrete random variable, X , has **probability mass function** (PMF), $p(\cdot)$, if $p(x) \geq 0$ and for all events A we have

$$P(X \in A) = \sum_{x \in A} p(x).$$

Definition. The **expected value** of a discrete random variable, X , is given by

$$E[X] := \sum_i x_i p(x_i).$$

Definition. The **variance** of any random variable, X , is defined as

$$\begin{aligned} \text{Var}(X) &:= E[(X - E[X])^2] \\ &= E[X^2] - E[X]^2. \end{aligned}$$

The Binomial Distribution

We say X has a binomial distribution, or $X \sim \text{Bin}(n, p)$, if

$$P(X = r) = \binom{n}{r} p^r (1 - p)^{n-r}.$$

For example, X might represent the number of heads in n independent coin tosses, where $p = P(\text{head})$. The mean and variance of the binomial distribution satisfy

$$\begin{aligned} E[X] &= np \\ \text{Var}(X) &= np(1 - p). \end{aligned}$$

A Financial Application

- Suppose a fund manager **outperforms** the market in a given year with probability p and that she **underperforms** the market with probability $1 - p$.
- She has a **track record** of 10 years and has outperformed the market in 8 of the 10 years.
- Moreover, performance in any one year is independent of performance in other years.

Question: How likely is a track record as good as this if the fund manager had no skill so that $p = 1/2$?

Answer: Let X be the number of outperforming years. Since the fund manager has no skill, $X \sim \text{Bin}(n = 10, p = 1/2)$ and

$$P(X \geq 8) = \sum_{r=8}^n \binom{n}{r} p^r (1-p)^{n-r}$$

Question: Suppose there are M fund managers? How well should the **best** one do over the 10-year period if none of them had any skill?

The Poisson Distribution

We say X has a **Poisson**(λ) distribution if

$$P(X = r) = \frac{\lambda^r e^{-\lambda}}{r!}.$$

$$E[X] = \lambda \text{ and } \text{Var}(X) = \lambda.$$

For example, the mean is calculated as

$$\begin{aligned} E[X] &= \sum_{r=0}^{\infty} r P(X = r) = \sum_{r=0}^{\infty} r \frac{\lambda^r e^{-\lambda}}{r!} = \sum_{r=1}^{\infty} r \frac{\lambda^r e^{-\lambda}}{r!} \\ &= \lambda \sum_{r=1}^{\infty} \frac{\lambda^{r-1} e^{-\lambda}}{(r-1)!} \\ &= \lambda \sum_{r=0}^{\infty} \frac{\lambda^r e^{-\lambda}}{r!} = \lambda. \end{aligned}$$

Bayes' Theorem

Let A and B be two events for which $P(B) \neq 0$. Then

$$\begin{aligned} P(A | B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(B | A)P(A)}{P(B)} \\ &= \frac{P(B | A)P(A)}{\sum_j P(B | A_j)P(A_j)} \end{aligned}$$

where the A_j 's form a partition of the sample-space.

An Example: Tossing Two Fair 6-Sided Dice

Y_2	6	7	8	9	10	11	12
	5	6	7	8	9	10	11
	4	5	6	7	8	9	10
	3	4	5	6	7	8	9
	2	3	4	5	6	7	8
	1	2	3	4	5	6	7
		1	2	3	4	5	6
		Y_1					

Table : $X = Y_1 + Y_2$

- Let Y_1 and Y_2 be the outcomes of tossing two fair dice **independently** of one another.
- Let $X := Y_1 + Y_2$. **Question:** What is $P(Y_1 \geq 4 | X \geq 8)$?

Continuous Random Variables

Definition. A continuous random variable, X , has **probability density function** (PDF), $f(\cdot)$, if $f(x) \geq 0$ and for all events A

$$P(X \in A) = \int_A f(y) \, dy.$$

The CDF and PDF are related by

$$F(x) = \int_{-\infty}^x f(y) \, dy.$$

It is often convenient to observe that

$$P\left(X \in \left(x - \frac{\epsilon}{2}, x + \frac{\epsilon}{2}\right)\right) \approx \epsilon f(x)$$

The Normal Distribution

We say X has a Normal distribution, or $X \sim \mathbf{N}(\mu, \sigma^2)$, if

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

The mean and variance of the normal distribution satisfy

$$\begin{aligned} \mathbf{E}[X] &= \mu \\ \text{Var}(X) &= \sigma^2. \end{aligned}$$

The Log-Normal Distribution

We say X has a log-normal distribution, or $X \sim \text{LN}(\mu, \sigma^2)$, if

$$\log(X) \sim N(\mu, \sigma^2).$$

The mean and variance of the log-normal distribution satisfy

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

The log-normal distribution plays a very important in financial applications.

Financial Engineering & Risk Management

Review of Conditional Expectations and Variances

M. Haugh G. Iyengar

Department of Industrial Engineering and Operations Research
Columbia University

Conditional Expectations and Variances

Let X and Y be two random variables.

The **conditional expectation identity** says

$$E[X] = E[E[X|Y]]$$

and the **conditional variance identity** says

$$\text{Var}(X) = \text{Var}(E[X|Y]) + E[\text{Var}(X|Y)].$$

Note that $E[X|Y]$ and $\text{Var}(X|Y)$ are both functions of Y and are therefore random variables themselves.

A Random Sum of Random Variables

Let $W = X_1 + X_2 + \dots + X_N$ where the X_i 's are IID with mean μ_x and variance σ_x^2 , and where N is also a random variable, independent of the X_i 's.

Question: What is $E[W]$?

Answer: The conditional expectation identity implies

$$\begin{aligned} E[W] &= E \left[E \left[\sum_{i=1}^N X_i \mid N \right] \right] \\ &= E[N\mu_x] = \mu_x E[N]. \end{aligned}$$

Question: What is $\text{Var}(W)$?

Answer: The conditional variance identity implies

$$\begin{aligned} \text{Var}(W) &= \text{Var}(E[W|N]) + E[\text{Var}(W|N)] \\ &= \text{Var}(\mu_x N) + E[N\sigma_x^2] \\ &= \mu_x^2 \text{Var}(N) + \sigma_x^2 E[N]. \end{aligned}$$

An Example: Chickens and Eggs

A hen lays N eggs where $N \sim \text{Poisson}(\lambda)$. Each egg hatches and yields a chicken with probability p , independently of the other eggs and N . Let K be the number of chickens.

Question: What is $E[K|N]$?

Answer: We can use **indicator functions** to answer this question.

In particular, can write $K = \sum_{i=1}^N 1_{H_i}$ where H_i is the event that the i^{th} egg hatches. Therefore

$$1_{H_i} = \begin{cases} 1, & \text{if } i^{\text{th}} \text{ egg hatches;} \\ 0, & \text{otherwise.} \end{cases}$$

Also clear that $E[1_{H_i}] = 1 \times p + 0 \times (1 - p) = p$ so that

$$E[K|N] = E\left[\sum_{i=1}^N 1_{H_i} \mid N\right] = \sum_{i=1}^N E[1_{H_i}] = Np.$$

Conditional expectation formula then gives $E[K] = E[E[K|N]] = E[Np] = \lambda p$.