

# Probability and Statistics Review

Jaehyuk Choi

Peking University HSBC Business School

2018-19 Module 1 (Fall 2018)

- Y1-M3: Stochastic Finance by Jaehyuk Choi [**required for Qfin MA**]
- Y1-M4: Derivative Pricing by Lei (Jack) Sun
- Y2-M1: Applied Stochastic Processes by Jaehyuk CHOI  
Application, Programming, Course project
- Y2-M1: Machine Learning for Finance by Jaehyuk CHOI  
(Topics in Quantitative Finance)
- Y2-M3: Numerical Methods and Analysis by Jake ZHAO
- Y2-M3: Bayesian Statistics by Qian CHEN

## New FinTech courses to come

- 2018-19 M3 (Y1): Blockchain
- 2018-19 M4 (Y1): Python & AI

- Random Variable (RV):  $U, X, Y, Z$
- Probability density function (PDF):  $f_X(x)$
- Cumulative density function (CDF):  $F_X(x) = \int f_X(x)dx$
- Standard deviation, variance:

$$\text{Var}(X) = E((X - \bar{X})^2) = E(X^2) - E(X)^2, \quad \sigma_X = \sqrt{\text{Var}(X)}$$

- (Centralized) Moments:  $M_k(X) = E((X - \bar{X})^k) = \int (x - \bar{X})^k f_X(x)dx$
- Moment generating function (MGF):  $M_X(t) = E(e^{tX})$

$$M_X(t) = 1 + tM_1 + \frac{t^2}{2!}M_2 + \cdots + \frac{t^k}{k!}M_k$$

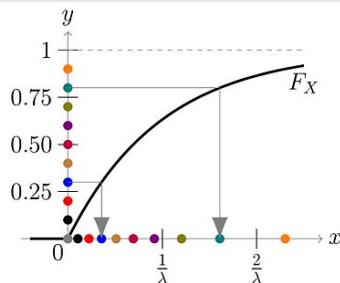
- Characteristic function (CF):  $\phi_X(t) = E(e^{itX}) + \cdots$
- Covariance:  $\text{Cov}(X, Y) = E((X - \bar{X})(Y - \bar{Y})) = E(XY) - E(X)E(Y)$
- Correlation:  $\rho(X, Y) = \text{Cov}(X, Y) / \sqrt{\text{Var}(X)\text{Var}(Y)} = \text{Cov}(X, Y) / (\sigma_X \sigma_Y)$

## Properties

- Support:  $[0, 1]$
- PDF:  $f(x) = 1$
- CDF:  $F(x) = x$
- Mean:  $E(U) = 1/2$
- Var:  $\text{Var}(U) = 1/12$

Uniform distribution is a fundamental RV which can be generated by computer. Once  $U$  is generated, any RV  $X$  is generated by **inverse transform sampling**

$$X = F_X^{-1}(U)$$

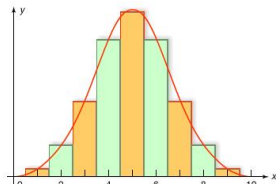
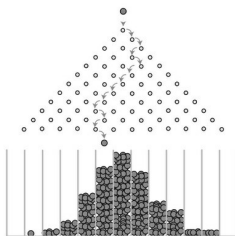


## Bernoulli distribution

- $P(X = 1) = p, \quad P(X = 0) = q = (1 - p)$
- $E(X) = p, \quad \text{Var}(X) = pq$

## Binomial distribution

- $Y = \sum_1^n X_k \sim N(n, p)$  for i.i.d. Bernoulli  $\{X_k\}$  with  $p$ .
- $P(Y = k) = \binom{n}{k} p^k q^{(n-k)}$
- $E(Y) = np, \quad \text{Var}(Y) = \sum_1^n \text{Var}(X_k) = npq$ .
- Approximated as normal dist. for large  $n$ :  $B(n, p) \approx N(np, npq)$



© 2005 Encyclopædia Britannica, Inc.

## Prob. Distribution: Event (default, arrival) at a constant rate $\lambda$

### Exponential distribution

- Distribution for the survival time or the interval between the events,  $T$
- PDF:  $f(t) = \lambda e^{-\lambda t}$ , CDF:  $F(t) = 1 - e^{-\lambda t}$
- $E(T) = 1/\lambda$ ,  $\text{Var}(T) = 1/\lambda^2$ .
- Memoryless: past events have no impact on the future!

### Poisson distribution (discrete)

- The number of occurrences  $X$  of a Poisson-type event in a unit time interval  $T = 1$
- PDF:  $P(X = k) = \lambda^k e^{-\lambda} / k!$
- $E(X) = \text{Var}(X) = \lambda$

### Gamma distribution

- The distribution of time  $X$  before the next  $k$  Poisson-type events occur
- $X \sim \Gamma(\alpha, \beta)$  where  $\alpha = k$ ,  $\beta = \lambda$ .
- PDF:  $f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$  for  $x \geq 0$  and  $\alpha, \beta > 0$ .
- $E(X) = \alpha/\beta$ ,  $\text{Var}(X) = \alpha/\beta^2$ .

# Normal (Gaussian) Distribution

- $X \sim N(\mu, \sigma^2), \quad Z \sim N(0, 1)$
- PDF:  $f_X(x) = \frac{1}{\sqrt{2\sigma^2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sigma} n\left(\frac{x-\mu}{\sigma}\right)$
- CDF:  $F_X(x) = N\left(\frac{x-\mu}{\sigma}\right)$
- MGF:  $M_X(x) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right), \quad M_k = \sigma^k(k-1)!!$  for even  $k$ .
- Skewness:  $s = M_3/\sigma^3 = 0$ , Kurtosis  $\kappa = M_4/\sigma^4 = 3$  (Ex-kurtosis: 0).

## Variations

- Multivariate normal distribution:  $(X_1, \dots, X_n)$
- Log-normal distribution:  $Y \sim e^{\mu + \sigma Z}$  for standard normal  $Z$ .

# Conditional Probability and Independence

## Conditional Probability

A probability of an event  $A$  given that an event  $B$  has occurred.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

## Independence

The two events  $A$  and  $B$  are (statistically) independent if  $P(A \cap B) = P(A)P(B)$ .  
Equivalently,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A) \quad \text{if } P(B) \neq 0$$
$$\text{and } P(B|A) = P(B) \quad \text{if } P(A) \neq 0$$

- Joint CDF:  $F_{X,Y}(x,y) = F_X(x)F_Y(y)$  ( $P(X \leq x, Y < y) = P(X \leq x)P(Y \leq y)$ )
- Joint PDF:  $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
- $E(XY) = E(X)E(Y)$ ,  $\text{Cov}(X,Y) = \rho(X,Y) = 0$ . However,  $\rho(X,Y) = 0$  does not imply independence.