Serik Sagitov, Chalmers Tekniska Högskola, December 21, 2012

An overview of the Probability Theory

1 Probability rules

Sample space: the set Ω of all possible outcomes in a random experiment.
Random events: $A, B \subset \Omega$, $A \cup B = \{A \text{ and } B\}$, $A \cap B = \{A \text{ or } B \text{ or both}\}$
Division rule: $\mathrm{P}(A) = \frac{\text{no. favorable outcomes}}{\text{total no. equally likely outcomes}}$
Addition rule: $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$
Complementary event probability: $\mathrm{P}(\bar{A}) = 1 - \mathrm{P}(A)$, $\bar{A} = \{A \text{ has not occurred}\}$
Conditional probability of A given B has occurred: $\mathrm{P}(A|B) = \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}$
Multiplication rule: $\mathrm{P}(A \cap B) = \mathrm{P}(A|B)\mathrm{P}(B)$

$$\boxed{\text{Independent events: } \mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)}$$

Law of Total Probability:
if $\{B_1, \ldots, B_k\}$ is a partition of Ω, then $\mathrm{P}(A) = \sum_{i=1}^{k} \mathrm{P}(B_i)\mathrm{P}(A|B_i)$
Bayes' Probability Law
prior probabilities $\mathrm{P}(B_i)$ and posterior probabilities $\mathrm{P}(B_i|A) = \frac{\mathrm{P}(A|B_i)\mathrm{P}(B_i)}{\mathrm{P}(A)}$

2 Random variables

Discrete random variable X: probability mass function (pmf) $f(x) = \mathrm{P}(X = x)$
Continuous random variable X: probab. density function (pdf) $f(x) \approx \frac{\mathrm{P}(x<X<x+\Delta)}{\Delta}$
Cumulative distribution function (cdf)

$$F(x) = \mathrm{P}(X \le x) = \sum_{y \le x} f(y) \text{ or } = \int_{y \le x} f(y)dy.$$

Mean (average or expected) value of X: $\mu = \mathrm{E}(X) = \sum_x xf(x)$ or $\mu = \int xf(x)dx$

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y), \quad \mathrm{E}(c \cdot X) = c \cdot \mathrm{E}(X)$$

Variance: $\sigma_X^2 = \mathrm{Var}(X) = \mathrm{E}((X-\mu)^2) = \mathrm{E}(X^2) - \mu^2$

$$\mathrm{E}(X^2) = \sum x^2 f(x) \text{ or } \mathrm{E}(X^2) = \int x^2 f(x)dx.$$

Standard deviation (SD): $\sigma_X = \sqrt{\mathrm{Var}(X)}$

$$\mathrm{Var}(c \cdot X) = c^2 \cdot \mathrm{Var}(X), \quad \sigma_{cX} = c \cdot \sigma_X$$

3 Special distributions

Discrete uniform distribution $X \sim \mathrm{dU}(N)$: $f(k) = \frac{1}{N}$, $1 \le k \le N$; $\mathrm{E}(X) = \frac{N+1}{2}$, $\mathrm{Var}(X) = \frac{N^2-1}{12}$
Uniform distribution $X \sim \mathrm{U}(a,b)$: $f(x) = \frac{1}{b-a}$, $a < x < b$, $\mathrm{E}(X) = \frac{a+b}{2}$, $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$
Binomial distribution $X \sim \mathrm{Bin}(n,p)$: $f(k) = \binom{n}{k} p^k q^{n-k}$, $0 \le k \le n$, $\mathrm{E}(X) = np$, $\mathrm{Var}(X) = npq$
Hypergeometric distribution $X \sim \mathrm{Hg}(N,n,p)$: $\mathrm{E}(X) = np$, $\mathrm{Var}(X) = npq(1-\frac{n-1}{N-1})$, finite population correction $\mathrm{FPC} = 1 - \frac{n-1}{N-1}$

$$f(k) = \frac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}}, \quad \max(0, n-Nq) \le k \le \min(n, Np)$$

Geometric distribution $X \sim \mathrm{Geom}(p)$: $f(k) = pq^{k-1}$, $k \ge 1$, $\mathrm{E}(X) = \frac{1}{p}$, $\mathrm{Var}(X) = \frac{q}{p^2}$
Exponential distribution $X \sim \mathrm{Exp}(\lambda)$: $f(x) = \lambda e^{-\lambda x}$, $x > 0$, $\mathrm{E}(X) = \sigma_X = \frac{1}{\lambda}$
Poisson distribution $X \sim \mathrm{Pois}(\lambda)$: $f(k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k \ge 0$, $\mathrm{E}(X) = \mathrm{Var}(X) = \lambda$
  $\mathrm{Bin}(n,p) \approx \mathrm{Pois}(np)$ if $n \ge 100$, $p \le 0.01$, $np \le 20$
Standard normal distribution $Z \sim \mathrm{N}(0,1)$: $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, $\mathrm{E}(Z) = 0$, $\mathrm{Var}(Z) = 1$,
  $\mathrm{P}(Z < z) = \Phi(z)$, $\mathrm{P}(|Z| > z) = 2(1 - \Phi(z))$
Normal distribution $X \sim \mathrm{N}(\mu, \sigma^2)$, $\frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$, $f(x) = \frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})$, $\mathrm{E}(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$
Central Limit Theorem (CLT):
  if $X_1, \ldots, X_n$ are IID, $\mathrm{E}(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$, then approximately $\bar{X} \overset{a}{\sim} \mathrm{N}(\mu, \frac{\sigma^2}{n})$ for large n
Normal approximations
  $\mathrm{Bin}(n,p) \approx \mathrm{N}(np, npq)$, $np \ge 5$, $nq \ge 5$
  $\mathrm{Pois}(\lambda) \approx \mathrm{N}(\lambda, \lambda)$, $\lambda \ge 5$
  $\mathrm{Hg}(N,n,p) \approx \mathrm{N}(np, npq\frac{N-n}{N-1})$, $np \ge 5$, $nq \ge 5$

4 Joint distributions

Joint pmf (pdf) of X and Y: $f_{X,Y}(x,y)$
Marginal distribution of X: $f_X(x) = \sum_y f_{X,Y}(x,y)$ or $\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$

$$\boxed{\text{Independent r.v. } X \text{ and } Y\text{: } f_{X,Y}(x,y) = f_X(x) f_Y(y)}$$

Conditional distribution of $(Y|X = x)$: $f_{Y|X}(y|x) = f_{X,Y}(x,y)/f_X(x)$
Conditional expectation $\mathrm{E}(Y|X)$ and $\mathrm{Var}(Y|X)$
  $\mathrm{E}(\mathrm{E}(Y|X)) = \mathrm{E}(Y)$
  $\mathrm{Var}(Y) = \mathrm{Var}(\mathrm{E}(Y|X)) + \mathrm{E}(\mathrm{Var}(Y|X))$
Addition rule for variance
  $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$
Covariance, a measure of association between X and Y
  $\sigma_{xy} = \mathrm{Cov}(X,Y) = \mathrm{E}(X - \mu_X)(Y - \mu_Y) = \mathrm{E}(XY) - \mathrm{E}(\mu_X)\mathrm{E}(\mu_Y)$
Correlation coefficient $\rho = \frac{\sigma_{xy}}{\sigma_X \sigma_Y}$, $-1 < \rho < 1$. Independent r.v. are always uncorrelated: $\rho = 0$
Multinomial distribution $(X_1, \ldots, X_r) \sim \mathrm{Mn}(n; p_1, \ldots, p_r)$
  $\mathrm{P}(X_1 = k_1, \ldots, X_r = k_r) = \binom{n}{k_1, \ldots, k_r} p_1^{k_1} \ldots p_r^{k_r}$, $\mathrm{Cov}(X_i, X_j) = -np_ip_j$
Bivariate normal distr $(X,Y) \sim \mathrm{N}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$. Marginal distributions : $X \sim \mathrm{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$. Conditional distribution of $(Y|X = x)$ is normal with mean $\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$, and variance $\sigma_y^2(1 - \rho^2)$