Probability Review

Arabin Kumar Dey

Assistant Professor

Department of Mathematics
Indian Institute of Technology Guwahati

24-th January, 2012

- Consider an experiment whose outcome is not known in advance.
- Sample space S is the set of all possible outcomes in the experiment.
- Any subset $A \subset S$ of the sample space is called an event.
- Complement of an event A, denoted by A^c , is the set of all possible outcomes not in A.
- Union of two events : $A \cup B$ consists of all outcomes that are either in A or in B or in both.
- Intersection of two events: $A \cap B$ (often denoted by AB) consists of all outcomes in both A and B.
- Mutually exclusive events: If $AB = \phi$, i.e. no outcomes in both A and B so that A and B can not both occur, we say they are mutually exclusive events.
- These definitions can naturally extend to more than a pair of events.

Axioms of probability

- $P(A)$ denotes a number associated with the probability of the event A.
- Three axioms of probability : a) $0 \leq P(A) \leq 1$
  b) $P(S) = 1$ with S, the sample space.
  c) For any sequence of mutually exclusive events $A_1, A_2, \cdots$
  $P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$ for $n = 1, 2, \cdots, \infty$
- Consequence : $P(A^c) = 1 - P(A)$ for any event A.

Conditional Probability

- For any two events A and B, we define

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

- Because for any event A, we have $A = AB \cup AB^c$, then

$$\begin{aligned} P(A) &= P(AB) + P(AB^c) \\ &= P(A|B)P(B) + P(A|B^c)P(B^c) \end{aligned}$$

- Let $B_1, B_2, \cdots, B_n$ be n mutually exclusive events whose union is the sample space S. Then we have the Law of Total Probability:

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Independence

- Two events A and B are said to be independent if

$$P(A|B) = P(A)$$

- As a consequence, we have A and B are two independent events if

$$P(AB) = P(A)P(B)$$

Random variables

- A random variable will be denoted by capital letters: X .
- X is a discrete random variable if it takes either a finite or at most countable number of possible values. Otherwise, it is continuous.
- Cumulative distribution function (cdf): $F(x) = P(X \leq x)$.
- Discrete: probability mass function (pmf) $p(x) = P(X = x)$.
- Continuous: probability density function (pdf) $f(x) = \frac{dF(x)}{dx}$
- It can be shown that: $F(x) = \int_{-\infty}^{x} f(z)dz$ and that

$$P(a - \frac{\epsilon}{2} < X < a + \frac{\epsilon}{2}) = \int_{a-\frac{\epsilon}{2}}^{a+\frac{\epsilon}{2}} f(z)dz \approx \epsilon f(a)$$

Multivariate Random Variable

- Sometimes called random vectors. Notation : (X, Y)
- Joint cdf: $F(x, y) = P(X \leq x, Y \leq y)$.
- Discrete pmf : $p(x, y) = P(X = x, Y = y)$.
- Continuous pdf : $f(x, y) = \frac{\delta^2 F(x,y)}{\delta x \delta y}$.
- For any set of real numbers C and D, we have

$$P(X \in C, Y \in D) = \int \int_{x \in C, y \in D} f(x, y) dx dy$$

- X and Y are said to be independent if

$$P(X \in C, Y \in D) = P(X \in C) P(Y \in D).$$

- Also, we have $P(X = x, Y = y) = P(X = x) P(Y = y)$ and $f(x, y) = f_X(x) f_Y(y)$ for two independent random variables X and Y .
- Note also that if X and Y are independent and so will $G(X)$ and $H(Y)$.

Expectation

- Discrete :
$$E(g(X)) = \sum_x g(x)p(x)$$
- Continuous :
$$E(g(X)) = \int_{-\infty}^{\infty} g(x)p(x)dx$$
- Linearity of expectation :
$$E(aX + b) = aE(X) + b$$
.
- Special cases:
  a) Mean of X : $\mu_X = E(X)$.
  b) Variance of X : $Var(X) = \sigma_X^2 = E[(X - \mu_X)^2]$.
  c) One can show that variance can also be expressed as
$$V(X) = E(X^2) - [E(X)]^2$$

Expectation - the multivariate case

- Discrete: $E(g(X,Y)) = \sum_x \sum_y g(x,y)p(x,y)$
- Continuous : $E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$
- Covariance of X and Y : $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$. One can show that you can also write this as $Cov(X,Y) = E(XY) - E(X)E(Y)$.
- Important property: $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$.
- If X and Y are independent, then $Cov(X,Y) = 0$ so that

$$Var(X+Y) = Var(X) + Var(Y).$$

  Note the converse is NOT always true: if $Cov(X,Y) = 0$, then X and Y are not necessarily independent.
- Correlation between X and Y :

$$corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

Inequalities and the laws of large numbers

- Markov's inequality: Suppose X takes only non-negative values. Then for any $a > 0$,

$$P(X > a) \leq \frac{E(X)}{a}.$$

- Chebyshev's inequality: Suppose X is a random variable with mean μ and variance σ^2. Then for any positive k,

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- The weak law of large numbers: Suppose $X_1, X_2, \cdots$ is a sequence of i.i.d. random variables with mean μ. Then for any $\epsilon > 0$,

$$P(|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu| > \epsilon) \to 0 \quad \text{as} \quad n \to \infty$$

- Strong law of large numbers: With certainty, the long-run average of a sequence of i.i.d. random variables will converge to its mean. That is,

The central limit theorem

Suppose $X_1, X_2, \cdots$ is a sequence of i.i.d. random variables with finite mean μ and finite variance σ^2. Then

$$\lim_{n\to\infty} P(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} < x) = \Phi(x)$$

This sample mean $\frac{X_1+X_2+\cdots+X_n}{n}$ is asymptotically normal. $\Phi(x)$ is the CDF of standard normal.

Discrete distributions

- Bernoulli
- Binomial
- Poisson
- Geometric
- Negative Binomial
- Hypergeometric

Continuous distributions

- Uniform
- Normal, log-Normal
- Exponential
- Gamma

Exponential Distribution

- X is Exponential with parameter λ if pdf is $f(x) = \lambda e^{-\lambda x}$, for $x > 0$.
- We write $X \sim Exp(\lambda)$.
- CDF: $F(x) = 1 - e^{-\lambda x}$
- Mean is $E(X) = \frac{1}{\lambda}$ and variance is $Var(X) = \frac{1}{\lambda^2}$.
- Some important properties:
  - Memoryless property : For all $s, t \geq 0$, we have
$$P(X > s + t | X > s) = P(X > t).$$
  - Multiplication of a constant: $X \sim Exp(\lambda)$ implies $cX \sim Exp(\frac{\lambda}{c})$ for any constant c.
  - Suppose $X_1, X_2, \cdots, X_n$ be n independent Exponential r.v.'s with respective parameters $\lambda_1, \lambda_2, \cdots, \lambda_n$. Then $M = \min\{X_1, X_2, \cdots, X_n\} \sim Exp(\sum_i \lambda_i)$, and the probability that X_j is the smallest is given by $\frac{\lambda_j}{\sum_j \lambda_j}$.

Gamma Random Variable

- A Gamma random variable with parameters $k > 0$ and $\lambda > 0$ has a pdf $f(x) = \frac{\lambda^k}{\Gamma(k)} e^{-\lambda x} x^{k-1} \qquad x > 0$
- We write $X \sim \text{Gamma}(k, \lambda)$ and the Gamma function is defined by $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$.
- Some properties: $\Gamma(z+1) = z\Gamma(z)$ and $\Gamma(n) = (n-1)!$, for any positive integer n.
- Mean is $E(X) = \frac{k}{\lambda}$ and variance is $Var(X) = \frac{k}{\lambda^2}$.
- If $X_1, \cdots, X_n$ are n i.i.d. Exponential r.v.'s with common parameter λ, then the sum $X = \sum_{i=1} X_i = X_1 + \cdots + X_n$ has a Gamma(n, λ) distribution.
- Proofs will be provided in lectures.

Normal Distribution

- X is said to be normally distributed with mean μ and variance σ^2 if its pdf has the form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

- We write $X \sim N(\mu, \sigma^2)$.
- In the case where $\mu = 0$ and $\sigma^2 = 1$, we have a standard Normal random variable Z.
- Note that $\frac{X-\mu}{\sigma} = Z \sim N(0, 1)$
- The cdf of a standatd Normal is usually written as

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$$

so that the cdf of a Normal can be evaluated from the standard Normal using $F(x) = \Phi(\frac{x-\mu}{\sigma})$.

Standard Normal density curve

- Percentile values for standard Normal distribution

lower tail area	0.9	0.95	0.975	0.99	0.995	0.999
horizontal axis value	1.282	1.645	1.960	2.326	2.576	3.090
upper tail area	0.1	0.05	0.025	0.01	0.005	0.001

- Horizontal axis values are called critical values.
- Tail areas (under the density curve) represent probabilities.
- Example: $P(Z > 1.960) = 0.025$ and $P(0 < Z < 1.960) = 1 - 0.5 - 0.025 = 0.475$.
  - $P(Z > 1.645) = ?$
  - $P(Z < -2.326 or > 2.326) = ?$
  - $P(Z < ?) = 0.90$ (i.e. what is the 90-th percentile?).
  - Also, evaluate the following probabilities: $P(-1 < Z < 1)$, $P(-2 < Z < 2)$, and $P(-3 < Z < 3)$.

Log-normal random variables

- The random variable X is said to have a log-normal distribution with parameters μ and σ^2 if $\log(X)$ is Normal(μ, σ^2).
- We write $X \sim LN(\mu, \sigma^2)$.
- It can be shown that its pdf has the form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}, \quad x > 0$$

- Mean is $E(X) = e^{\mu + \frac{\sigma^2}{2}}$ and variance is $Var(X) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$.
- Caution: Do not make the mistake that the mean and variance of a log-normal random variable are μ and σ^2 .

Homogeneous Poisson process

- Let $N(t)$ be the number of events that occur in a time interval $[0, t]$. Then $\{N(t), t \geq 0\}$ is a poisson process with rate $\lambda > 0$ if
  - $N(0) = 0$
  - It has independent increments, i.e. number of events occurring in disjoint time intervals are independent.
  - It has stationary increments, i.e. distribution of the number of events that occur in a given interval depends only on the length of the interval and not on its location.
  - It has unit jumps, i.e. $\lim_{h \to 0} \frac{P(N(h)=1)}{h} = \lambda$ and $\lim_{h \to 0} \frac{P(N(h) \geq 2)}{h} = 0$
- As a consequence of this definition, it can be shown that "the number of events in any interval of lenght t has a Poisson distribution with parameter λt". That is, for all $s, t \geq 0, \quad n = 0, 1, \cdots,$

$$P(N(t+s) - N(s) = n) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

Interarrival times

- Let T_1 denote the time of the first event. For $n > 1$, let T_n denote the elapsed time between the (n - 1)th and the nth events.
- $T_1, T_2, \cdots$ are called the interarrival times.
- Result: The interarrival times $T_1, T_2, \cdots$ are i.i.d. Exponential random variables with parameter λ. (Proof in
- The arrival time of the nth event (or the waiting time until the nth event) is the sum of the interarrival times, i.e.

$$S_n = T_1 + T_2 + \cdots + T_n.$$

- Obviously, the arrival time has a Gamma distribution (what are the parameters?).

Conditional expectation and conditional variance

- Law of iterated expectation: For any random variables X and Y , we have

$$E(E(X|Y)) = E(X)$$

- The conditional variance formula:

$$V(X) = E(V(X|Y)) + V(E(X|Y))$$

- These formulas are very important in many places, e.g. acturial.

Here is another job interview question. You die and are presented with three doors. One of them leads to heaven, one leads to one day in purgatory, and one leads to two days in purgatory. After your stay in purgatory is over, you go back to the doors and pick again, but the doors are reshuffled each time you come back, so you must in fact choose a door at random each time. How long is your expected stay in purgatory?