Compact course notes

Stat 240/230, Fall 2010

Probability

Professor: X.Zou, K.Moshksar
transcribed by: J. Lazovskis
University of Waterloo
December 13, 2010

Contents

File I

Stat 240

1 Introduction to Probability

1.1 Base definitions

Probability is a quantitative way to manipulate uncertainty..

Definition 1.1.1. A sample space is the connection of all possible results for an experiment. This is denoted by Ω. All elements in a sample space must be unique.

Remark 1.1.2. A sample space may contain more than all the possible results.

Definition 1.1.3. For a given sample space, an event A is any subset of Ω.

Operations of events. For $A, B \in \Omega$:

$A \cup B \equiv A + B$	A or B
$A \cap B \equiv A \cdot B$	A and B
$A - B$	in A but not in B
$A^C \equiv \Omega - A$	

The set F is the collection of all the events we are interested in for an experiment.

Definition 1.1.4. F is termed a σ-algebra if:

1. For any $A \in F$, $A^C \in F$
2. If $A_1, A_2, \ldots, A_n, \cdots \in F$, then $\displaystyle\bigcup_{i=1}^{\infty} A_i \in F$

Remark 1.1.5. (Ω, F) is termed a measurable space. For an Ω, there are two trivial σ-algebras: $F = \{\emptyset, \Omega\}$. If $|\Omega|$ is finite, then the power of Ω is used as the default σ-algebra. A power set of Ω is the set of all subsets of Ω.

1.2 Conditional probability

Definition 1.2.1. Given events $A, B \in \Omega$, the probability of event A occurring, given that event B has occurred, is

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

2 Random variables

2.1 Measurable functions

Definition 2.1.1. Given a probability model (Ω, F, P), a real-valued function X on Ω is termed a measurable function if it is a simple function or a limit of a sequence of simple functions.

Theorem 2.1.2. f is measurable $\iff f^{-1}((-\infty, a]) \in F$ for all $a \in \mathbb{R}$

$$= \{\omega \,\big|\, f(\omega) \leqslant a\}$$

Definition 2.1.3. A measurable function is termed a random variable.

Definition 2.1.4. The mean value operator E denotes the mean value of a random variable X. It is a linear operator.

3 Distributions

Definition 3.0.1. Given a random variable X, define
the cumulative distribution function (cdf): $F_X(x)$
the probability density function (pdf): $f_X(x) = \frac{d}{dx} F_X(x)$

3.1 Discrete distributions

Definition 3.1.1. A distribution is termed uniformly discrete if the distribution function $m(\omega)$ is defined to be $1/n$ for all $\omega \in S$, where S is the sample space of finite size n.

Definition 3.1.2. A Bernoulli trials process is a sequence of n chance experiments such that:
1. Each experiment has two possible outcomes
2. The probability of outcome 1, p, is the same for each experiment, and is not affected by previous experiments. The probability of outcome 2 is then $q = 1 - p$.

Definition 3.1.3. Given n Bernoulli trials with probability p of the first outcome, the probability of exactly j outcomes 1 is

$$b(n, p, j) = \binom{n}{j} n^j q^{n-j}$$

This is also termed the binomial distribution

Definition 3.1.4. Given a Bernoulli trials process repeated an infinite number of times, let T be the number of trials up to and including the first occurrence of outcome 1. Then

$$P(T = n) = q^{n-1} p$$

This is termed geometric distribution for the random variable T.

Theorem 3.1.5. Given a random variable T with geometric distribution and $r, s \geqslant 0$,

$$P(T > r + s \mid T > r) = \frac{P(T > r + s)}{P(T > r)} = q^s$$

Definition 3.1.6. Given n Bernoulli trials, let X be the random variable that represents the number of experiments up to and including the kth result of outcome 1. Then X is said to have negative binomial distribution. Note that for $k = 1$, we have geometric distribution.

Definition 3.1.7. Suppose there is a set of N balls, with k red balls and $N - k$ blue balls. When n balls are chosen, X is the number of red balls in the chosen sample. Then X has hypergeometric distribution, which is given by

$$h(N, k, n, x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

3.2 Continuous distributions

Definition 3.2.1. Suppose (Ω, F, P) is a probability model. Let X be a random variable $X : \Omega \to \mathbb{R}$. Then X is termed continuous if $f(x) = \lim_{\Delta x \to 0^+} \left[\frac{P(x \leqslant X \leqslant x + \Delta x)}{\Delta x} \right]$ exists for all $x \in \mathbb{R}$.

Remark 3.2.2. Then for all x, $P(x) = 0$.

Theorem 3.2.3. For a continuous random variable X, the expectation and variance is described by

$$\mu = E[x] = \int_{-\infty}^{\infty} x \cdot f(x) dx$$
$$\sigma^2 = V(x) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Theorem 3.2.4. If X is a continuous random variable and E is any event with positive probability, then define the continuous conditional probability of X given E to be $f(x|E) = \begin{cases} \frac{f(x)}{P(E)} & \text{if } x \in E \\ 0 & \text{else} \end{cases}$

4 Multiple random variables

4.1 Independent random variables

Definition 4.1.1. If $X \perp Y$, then $P(X \leqslant x, Y \leqslant y) = P(X \leqslant x)P(Y \leqslant y)$. This implies that $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for all Borel sets A, B.

Definition 4.1.2. If $X \perp Y$, then $E[g(x)h(y)] = E[g(x)]E[h(y)]$ for any measurable functions g, h on $\mathbb{R}$.

4.2 Variance and covariance

Definition 4.2.1. Let X be a numerically-valued random variable with expected value $\mu = E[x]$. Then the variance of X is given by: $V(X) = E[(x-\mu)^2]$

Note that V is not a linear operator.

Definition 4.2.2. Let X, Y be two random variables. Then the covariance of X and Y is a measure of the interdependency of X and Y. It is given by:

$$\begin{aligned} Cov(X,Y) &= E\left[(x-\mu_x)(y-\mu_y)\right] \\ &= E\left[(x-E[x])(y-E[y])\right] \\ &= E[xy] - E[x]E[y] \end{aligned}$$

Remark 4.2.3. If $Cov(X,Y) > 0$, then Y tends to increase as X increases. If $Cov(X,Y) < 0$, then Y tends to decrease as X increases.

Remark 4.2.4. $Cov(X,X) = V(X) = E[x^2] - E[x]^2$

Remark 4.2.5. If X and Y are independent, then $Cov(X,Y) = 0$. However, $Cov(X,Y) = 0 \not\Longrightarrow X \perp Y$.

Proposition 4.2.6. $Cov(X,Y) = Cov(Y,X)$

4.3 Functions of a random variable

Theorem 4.3.1. Let X be a continuous random variable. Suppose $\varphi(x)$ is a strictly increasing function on the range of X. Let $Y = \varphi(x)$. Suppose X, Y have cdf's F_X, F_Y. Then

$F_Y(y) = F_X(\varphi^{-1}(y))$

If $\phi(x)$ is strictly decreasing on the range of X, then

$F_Y(y) = 1 - F_X(\varphi^{-1}(y))$

If $\phi(x)$ is neither strictly increasing nor strictly decreasing on the range of X, then

$F_Y(y) = F_X(\varphi_+^{-1}(y)) - F_X(\varphi_-^{-1}(y))$

where $\varphi_+^{-1}(y)$ describes φ^{-1} where it is strictly increasing, and $\varphi_-^{-1}(y)$ describes φ^{-1} where it is strictly decreasing.

5 CDF & PDF

5.1 Cumulative distribution function

5.2 Probability density function

If X is a continuous random variable with density function $f(x)$, and E is an event with positive probability, then the conditional density function is given by

$$f(x|E) = \begin{cases} f(x)/P(E) & x \in E \\ 0 & else \end{cases}$$

File II
Stat 230

6 Event arithmetic

Definition 6.1. Two events A, B are said to be mutually exclusive when $AB = \emptyset$. In this case, either one of two events will occur, or neither will occur.

If A, B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$.

Theorem 6.2. [de Morgan's Laws]
Let $I \subset \mathbb{N}$. Then for all events or sets A_α for $\alpha \in I$,

- $\left(\bigcup_{\alpha\in I} A_\alpha\right)^c = \bigcap_{\alpha\in I} A_\alpha^c$
- $\left(\bigcap_{\alpha\in I} A_\alpha\right)^c = \bigcup_{\alpha\in I} A_\alpha^c$

Theorem 6.3. [Bayes' theorem]
Given events A, B, the conditional probability of A given B may be expressed as

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A^c)P(A^c) + P(B|A)P(A)}$$

7 Distributions

For a discrete random variable X, $f_X(x)$ is the mass probability function.
For a continuous random variable X, $f_X(x)$ is the probability density function.
For a compound random variable $X = X_1 + \cdots + X_k$, $f_{X_1,\ldots,X_k}(x_1, \ldots, x_k)$ is the joint probability function.

Remark 7.1. The three assumptions made when using a Poisson process distribution are:

- Independence
- Individuality
- Homogeneity

Definition 7.0.1. For a discrete random variable X with probability function $f(x)$, the moment generating function of X is defined to be

$$M(t) = E(e^{tX}) = \textstyle\sum_x e^{tx} f(x)$$

8 Distributions

Notation	Function	Mean	Variance	Moment Gen Func
Binomial(n,p,x) $0<p<1, q=1-p$	$\binom{n}{x} p^x(1-p)^{n-x}$	np	$np(1-p)$	$(pe^t+q)^n$
Bernoulli(p) $0<p<1, q=1-p$	$p(1-p)^{x-1}$	p	$p(1-p)$	(pe^t+q)
Negative binomial(k,p,x) $0<p<1, q=1-p$	$\binom{x-1}{k-1} p^k q^{x-k}$	$\frac{kq}{p}$	$\frac{kq}{p^2}$	$\left(\frac{p}{1-qe^t}\right)^k$
Geometric(x) $0<p<1, q=1-p$	pq^x	$\frac{q}{p}$	$\frac{q}{p^2}$	$\left(\frac{p}{1-qe^t}\right)$
Hypergeometric(N,k,n,x) $k<N, n<N$	$\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$	$\frac{nk}{N}$	$\frac{nk}{N}\left(1-\frac{k}{N}\right)\frac{N-n}{N-1}$	-
Poisson(μ,x) $\mu>0$	$\frac{e^{-\mu}\mu^x}{x!}$	μ	μ	$e^{\mu(e^t-1)}$
Notation	**PDF**	**Mean**	**Variance**	**MGF**
Uniform(a,b) $a<b$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt}-e^{at}}{(b-a)t}$
Exponential(λ) $\lambda>0$	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\left(\frac{1}{\lambda}\right)^2$	$\frac{\lambda}{\lambda-t}$
Normal(μ,σ^2) $-\infty<\mu<\infty, \sigma^2>0$	$\frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$	μ	σ^2	$exp\left\{\mu t+\frac{\sigma^2t^2}{2}\right\}$

Definition 8.1. A <u>Bernoulli trials process</u> is a sequence of n chance experiments such that

1. Each experiment has two outcomes, success and failure
2. Probability p for success stays constant for successive experiments, which are independent

Definition 8.2. The <u>binomial distribution</u> is the probability of exactly x successes in n Bernoulli trials.

Definition 8.3. If X is the number of trials until the kth success of a Bernoulli trials process, then X has <u>negative binomial</u> distribution.

Definition 8.4. If X is the number of trials (possibly infinite) until the 1st success of a Bernoulli trials process, then X has <u>geometric</u> distribution.

Definition 8.5. Suppose in a set of N items, k are type A, $N-k$ are type B. If n items are chosen at random, the probability that exactly x will be type A has <u>hypergeometric distribution</u>.

Definition 8.6. If a certain event occurs randomly with probability p over a period of time t at a constant rate of λ per some time interval, then it has <u>Poisson distribution</u>.