MATH 323: Study Sheet

Ryan Ordille

Exam date: 17 April, 2012

1 Before random variables

1.1 Basic axioms and theorems

Principle axioms:

1. $P(E) \geq 0$
2. $P(S) = 1$
3. For disjoint events E_1, E_2: $P(E_1 \cup E_2) = P(E_1) + P(E_2)$.

Principle theorems:

1. $P(\emptyset) = 0$
2. $P(E^c) = 1 - P(E)$
3. $P(E \cap F^c) = P(E) - P(E \cap F)$
4. $E \subseteq F \Rightarrow P(E) \leq P(F)$
5. For *any* events E, F: $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

If all outcomes are equally likely, i.e. $P(E_i) = P(E_j)$ for $i \neq j$, then:

$$P(E) = \frac{\text{Number of ways } E \text{ can occur}}{\text{total number of possible outcomes}}$$

1.1.1 Conditional probability

Definition:

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

Law of Total Probability: if $B_1, B_2, \ldots, B_m$ form a partition of S and $B_i \cap B_j = \emptyset$ for $i \neq j$, then:

$$P(A) = \sum_{i=1}^{m} P(A \mid B_i) P(B_i)$$

Bayes' Theorem: (assuming same premises as above)

$$P(B_k \mid A) = \frac{P(A \mid B_k)P(B_k)}{P(A)}$$

Conditional expectation:

$$P(A \cap B) = P(B \mid A)P(A) = P(A \mid B)P(B)$$

1.1.2 Independence

$$A \perp B \Leftrightarrow P(B \mid A) = P(B) \text{ or } P(A \mid B) = P(A)$$
$$A \perp B \Leftrightarrow P(A \cap B) = P(A)P(B)$$

2 Random variable distributions

2.1 C.d.f. and p.f.

Cumulative distribution function: (c.d.f.) $F_X(x) = P(X \leq x)$

Probability function (p.f.) $P_X(x) = P(X = x)$ for discrete r.v.s

2.2 Named distribution p.f.s

Bernoulli $X \sim Bern(p)$: where $P(X = 1) = p$ "success" and $P(X = 0) = 1 - p$ "failure"

Binomial $X \sim Bin(n, p)$: "number of successes in n trials" (for $x = 0, 1, \ldots$)

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

Poisson $X \sim Po(\lambda)$: "large n and small p binomial" where $np = \lambda$

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

Hypergeometric $X \sim HG(N, a, n)$: "number a of successes in n draws of N objects"

$$P(X = x) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$$

Geometric $X \sim Geom(p)$: "trial number of first success" (for $x = 1, 2, \ldots$)

$$P(X = x) = (1 - p)^{x-1} p$$

3 Expected values and variances

3.1 Definitions

Expected Value: "centre" of distribution, mean

$$E(X) = \mu = \sum_x x P_X(x)$$

$E(cX) = cE(X)$, $E(X^2) = \sum x^2 P(X = x)$, $E(XY) \neq E(X)E(Y)$, $E(\sum X_i) = \sum E(X_i)$

Variance: "spread" of the distribution

$$Var(X) = \sigma^2 = E((X - \mu_X)^2) = E(X^2) - \mu^2$$

$Var(cX) = c^2 Var(X)$

Standard deviation: $\sqrt{Var(X)} = \sqrt{\sigma^2} = \sigma$

3.2 Named distribution expected values and variances

Binomial $X \sim Bin(n, p)$: $E(X) = np$, $Var(X) = np(1 - p)$

Bernoulli $X \sim Bernoulli(p)$: equivalent to $Binom(1, p)$

Poisson $X \sim Po(\lambda)$: $E(X) = Var(X) = \lambda$

Geometric $X \sim Geom(p)$: $E(X) = \frac{1}{p}$, $Var(X) = \frac{1-p}{p^2}$

4 Probability density function p.d.f

4.1 Definitions

As $P(X = x) = 0$ for all x in a continuous distribution, our definition of the probability function is not enough for continuous distributions.

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(y)\, dy$$

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx$$

$$Var(X) = \int_{-\infty}^{\infty} x^2\, f_X(x)\, dx - \left(\int_{-\infty}^{\infty} x\, f_X(x)\, dx \right)^2$$

4.2 P.d.f.s of named distributions

4.2.1 Uniform $X \sim U(a, b)$:

$$f_X(x) = \int_a^x \frac{1}{b-a}\, dx$$

for $a \leq x \leq b$, and $f_X(x) = 0$ elsewhere. The c.d.f grows linearly, with $F_X(x) = \frac{x-a}{b-a}$ where $a \leq x \leq b$.

4.2.2 Exponential $X \sim Exp(\beta)$:

$$f_X(x) = \frac{1}{\beta} e^{-x/\beta} = \lambda e^{\lambda x}$$

for $x \geq 0$, with $f_X(x) = 0$ elsewhere.

$$F_X(x) = 1 - e^{-x/\beta}$$

$E(X) = \mu = \beta$, $Var(X) = \sigma^2 = \beta^2$.

Memoryless property: if $X \sim Exp(\beta)$, then $P(x \leq X < x + h \mid X \geq x) = P(0 \leq x < h)$.

4.2.3 Gamma $X \sim Gamma(\alpha, \beta)$:

Gamma function: let $\alpha > 0$:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\, e^{-x}\, dx$$

where $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$.

Gamma distribution:

$$f_X(x) = \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \text{ for } x \geq 0$$

with $f_X(x) = 0$ elsewhere.

$E(X) = \alpha\beta$, $Var(X) = \alpha\beta^2$.

$$F_X(x) = \int_0^x \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1}\, dy \text{ for } x > 0$$

4.2.4 Chi-Square $X \sim \chi^2_\nu$:

$$f_X(x) = \frac{1}{\Gamma(\nu/2)2^{\nu/2}} x^{\nu/2-1} \text{ for } x \geq 0$$

with $f_X(x) = 0$ elsewhere.

4.2.5 Gaussian/Normal $X \sim N(\mu, \sigma^2)$:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \ \forall x$$

$E(X) = \mu$ and $Var(X) = \sigma^2$.

4.3 Standardizing distributions

4.3.1 Any r.v.:

To standardize any r.v. X with mean μ and standard deviation σ:

$$Y = \frac{X - \mu}{\sigma}$$

where $E(Y) = 0$ and $Var(Y) = 1$.

4.3.2 For $X \sim N(\mu, \sigma^2)$:

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

where $Z \sim N(0, 1)$ is a *standard normal* random variable.

5 Moment generating functions and transformations

5.1 Moment generating function m.g.f.

$$M_X(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx}\ f_X(x)\ dx = \sum_x e^{tx}\ P(X = x)$$

$E(X^k) = M_X^{(k)}(0)$

5.2 Transformations

5.2.1 Transforming p.d.f.s

Let $y = g(x)$ be either strictly decreasing or strictly increasing, X be continuous with p.d.f. f_X, and $Y = g(X)$. Then, the p.d.f. of Y is:

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dx}{dy} \right|$$

5.2.2 Probability integral transformation

Let X be a continuous random variable with a strictly increasing cumulative distribution function F_X, and $Y = F_X(X)$. Then, $Y \sim U(0,1)$.

6 Joint distributions

6.1 Definitions

Joint c.d.f: $F_{X,Y}(x,y) = P(X \leq x \cap Y \leq y) = P(X \leq x, Y \leq y)$

Marginal c.d.f.:

$$F_X = \lim_{y \to +\infty} F_{X,Y}(x,y)$$

Joint p.d.f.: (continuous)

$$F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u,v)\ du\ dv$$

(discrete)

$$F_{X,Y}(x,y) = \sum_{v \leq y} \sum_{u \leq x} P_{X,Y}(u,v)$$

Use the Fundamental Theorem of Calculus to extract $f_{X,Y}(x,y)$.

Joint probability function: $P_{X,Y}(x,y) = P(X = x, Y = y)$

Marginal p.d.f.: integrate out the variable you wish to get rid of:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dx$$

6.2 Conditional distributions

6.2.1 Conditional probability function

Definition: given the joint probability function of (X,Y), $P_{X,Y}(x,y) = P(X = x, Y = y)$:

$$P(Y = y \mid X = x) = \frac{P_{X,Y}(x,y)}{P_X(x)} = F_{Y \mid X \leq x}(y \mid X \leq x)$$

6.2.2 Conditional p.d.f.

$$f_{Y \mid X=x}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$$F_{Y \mid X=x}(y \mid x) = \int_{-\infty}^{y} f_{Y \mid X=x}(u \mid x)\, du$$

$$E(Y \mid X = x) = \int_{-\infty}^{\infty} y\, f_{Y \mid X=x}(y \mid x)\, dy$$

6.2.3 Law of Total Probability for Random Variables

Discrete:

$$P(Y = y) = \sum_x P(Y = y \mid X = x)P(X = x) = \sum_x P_{Y \mid X=x} P_X(x)$$

Continuous:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y \mid X=x}(y \mid x)\, f_X(x)\, dx$$

7 Covariance and correlation

7.1 Covariance

Continuous:

$$E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\, dx\, dy$$

Discrete:

$$E(g(X,Y)) = \sum_y \sum_x g(x,y) P_{X,Y}(x,y)$$

Covariance: when $g(X,Y) = (X - \mu_X)(Y - \mu_Y)$, the *covariance* between X and Y is $E(g(X,Y))$, denoted $Cov(X,Y)$ or $\sigma_{X,Y}$.

The covariance is a measure about the means of X and Y – if $Cov(X,Y) > 0$, then as X increases, Y also increases (same for decreasing), while if $Cov(X,Y) < 0$, then as X increases, Y decreases (and vice versa).

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

$Cov(X,Y)$'s magnitude depends on the scale of measurement (i.e. $Cov(aX, bY) \neq Cov(X,Y)$).

7.2 Correlation

Correlation:

$$|Corr(X,Y)| = |\rho(X,Y)| = \left|\frac{\sigma_{XY}}{\sigma_X \sigma_Y}\right| = \left|\frac{Cov(aX,bY)}{\sqrt{Var(aX)Var(bY)}}\right| = \frac{|ab|Cov(X,Y)}{|ab|\sqrt{Var(X)Var(Y)}}$$

$Corr(X,Y)$ has the same sign as $Cov(X,Y)$. The correlation between X and Y is a measure of the linear dependence between them, *not* the general dependence between them.

7.3 Covariance and correlation

$$Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X,Y)$$

If $Cov(X,Y) = 0$, then $Var(X+Y) = Var(X) + Var(Y)$, and, in general:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

if X_i, X_j are uncorrelated for $i \neq j$.

8 Independence between random variables

The random variables $X_1, X_2, \ldots, X_n$ are said to be independent if and only if

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_n}(x_n)$$

If X and Y are independent, then $Cov(X,Y) = 0$ and $E(XY) = E(X)E(Y)$. The converse is not always true, however.

9 Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed with mean μ and variance σ^2. Then,

$$P\left(\left(\frac{S_n - n\mu}{\sqrt{n}\sigma}\right) \leq x\right) \to P(Z \leq x)\ \forall x \text{ as } n \to \infty$$