4113 Review (w/o expls)

Sections 2.1-2.3

Events

A = "[some declarative sentence w/a truth value]." ← How we think about events

A = {outcomes where that sentence is true} ← How we work with events

Translation:

sentences	sets
not-A	A^c
A & B	$A \cap B$
A or B	$A \cup B$
$A \Rightarrow B$	$A \subseteq B$

Ω = "[assumptions of model]"
= {all possible outcomes}

Probability model: (Ω, P)

P: function mapping A, B to $P(A|B)$

ALL PROBABILITIES ARE RELATIVE (or conditional)

Shorthand: $P(A) = P(A|\Omega)$, $P(A|B) = P(A|B \,\&\, \Omega)$

Intuitive version of axioms:

Ax. I: $0 \le P(A|B) \le 1$

Ax. II: If $B \subseteq A$, then $P(A|B) = 1$

Ax. III: If $C \subseteq (A \cap B)^c$, then

$$P(A \cup B | C) = P(A|C) + P(B|C)$$ (addition rule)

Ax. IV: If $P(A|C) > 0$, then

$$P(A \cap B | C) = P(A|C)\,P(B|A \cap C)$$ (mult. rule)

Ax. V: If $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$$ (continuity rule)

Classical version of axioms:

Ax. 1: $0 \le P(A) \le 1$

Ax. 2: $P(\Omega) = 1$

Ax. 3: If $A_1, A_2, \ldots$ are mut. excl., then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$ (gives add. rule and cont. rule)

Defn: If $P(B) > 0$, then

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$ (gives mult. rule)

4113 Review

Ch. 1

$n!$: # of ways to permute n objects

$\frac{n!}{(n-r)!}$: # of ways to choose r from n, when order matters

$\binom{n}{r}$: # of ways to choose r from n, when order doesn't matter

$\binom{n}{n_1, \ldots, n_r}$: # of ways to break n objects into r distinct groups of sizes $n_1, \ldots, n_r$ (requires $n_1 + \cdots + n_r = n$)

Properties

$0! = 1$

$n! = n \cdot (n-1)!$ for integers $n > 1$

If n, r are integers,

$$\binom{n}{r} = \begin{cases} \frac{n!}{(n-r)!\,r!} & \text{if } 0 \le r \le n, \\ 0 & \text{o.w.} \end{cases}$$

$$\binom{n}{n_1, \ldots, n_r} = \frac{n!}{n_1! \cdots n_r!}$$

$$\binom{n}{n-r} = \binom{n}{r}$$

$$\binom{n}{0} = \binom{n}{n} = 1 \,, \quad \binom{n}{1} = \binom{n}{n-1} = n$$

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r} \qquad \text{(binomial thm)}$$

$$\binom{n}{n_1, n_2} = \binom{n}{n_1} = \binom{n}{n_2}$$

4113 Review

Sections 2.4–2.5

Prop. 2.4.1 $P(A^c) = 1 - P(A)$

Prop. 2.4.2 $A \subseteq B \Rightarrow P(A) \le P(B)$

Prop. 2.4.3

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

(inclusion-exclusion)

Can do w/ more than 2 events:

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) = \sum_{j=1}^{4} P(A_j)$$

stop here and its too big

$$- \sum_{i<j} P(A_i \cap A_j)$$

stop here and its too small

$$+ \sum_{i<j<k} P(A_i \cap A_j \cap A_k)$$

$$- P(A_1 \cap A_2 \cap A_3 \cap A_4)$$

If Ω is finite and all outcomes are equ. likely, then

$$P(A) = \frac{|A|}{|\Omega|}$$

# of elements in A

4113 Review

Chapter 3

Defn of cond. prob: $P(B|A) = \dfrac{P(A \cap B)}{P(A)}$

Mult. rule: $P(A \cap B) = P(A)\, P(B|A)$

Extensions of mult. rule:

$P(A \cap B \cap C) = P(A)\, P(B|A)\, P(C|A \cap B)$ (can do more than 3 also)

$P(A \cap B | C) = P(A|C)\, P(B|A \cap C)$

Law of total prob.: If $B_1, \dots, B_n$ is a partition, then

$$P(A) = \sum_{j=1}^{n} P(B_j)\, P(A|B_j)$$

Special case: $P(A) = P(B)P(A|B) + P(B^c)P(A|B^c)$

Defn of "independent":

A, B indep. if $P(A \cap B) = P(A)P(B)$

Alternative conditions:

- A, B indep. if $P(A|B) = P(A)$
- A, B indep. if $P(B|A) = P(B)$

$A_1, \dots, A_n$ indep. if, for every subset $J \subseteq \{1, \dots, n\}$,

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j).$$

$A_1, A_2, \dots$ indep. if, for every n, $A_1, \dots, A_n$ indep.

Combos of different indep. events are indep. E.g., if $A_1, \dots, A_7$ are indep, and

$$B_1 = A_1 \cap (A_4 \cap A_6)^c$$
$$B_2 = A_2 \cup A_7$$

then B_1 and B_2 are indep.

The function $P(\cdot \mid C)$ satisfies the axioms.

$\therefore$ can add conditioning to any thm. E.g.,

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

becomes

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

A and B are conditionally indep. given C if

$$P(A \cap B \mid C) = P(A \mid C)\,P(B \mid C).$$

extends to multiple events as above

OR
- $P(A \mid B \cap C) = P(A \mid C)$
- $P(B \mid A \cap C) = P(B \mid C)$

4113 Review (no expls)

Chapter 4

Random Variables (Sections 4.1, 4.10)

a quantity whose value is uncertain

Expl: X = "the number of water molecules in my body"

describe by a noun phrase denoting a number

Formal defn:
A random variable is a function $X : \Omega \to \mathbb{R}$

Shorthand:

$$P(X \in A) = P(\{X \in A\}) = P(\underbrace{\{\omega \in \Omega \mid X(\omega) \in A\}}_{\text{also called } X^{-1}(A)})$$

distribution function: $F_X(x) = P(X \le x)$ defined for all $x \in \mathbb{R}$

Notation:

$$F(\infty) = \lim_{s \to \infty} F(s)$$

$$F(-\infty) = \lim_{s \to -\infty} F(s)$$

$$F(x-) = \lim_{s \to x^-} F(s)$$

Properties

- F is nondecreasing and right-continuous
- $F(\infty) = 1$, $F(-\infty) = 0$
- $P(X < x) = F(x-)$
- $P(X = x) = F(x) - F(x-)$

Discrete r.v.s (Sections 4.2–4.5, 4.9)

X is **discrete** if $\underbrace{R(X)}_{\text{range of } X}$ is countable

mass function: $p_X(x) = P(X = x)$

defined for all $x \in \mathbb{R}$

expected value: $E[X] = \sum_{x \in R(X)} x \, P(X = x)$

indicator r.v.: $\underbrace{1_A}_{\text{a r.v.}} = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{if } A \text{ is false} \end{cases}$

an event

$$\boxed{E[1_A] = P(A)}$$

alternate formula: $E[X] = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\})$

Properties

- $E[g(X)] = \sum_{x \in R(X)} g(x) P(X = x)$

true even when not discrete

- $E[cX] = c E[X]$
- $E[X + Y] = E[X] + E[Y]$

variance: $Var(X) = E[(X - E[X])^2]$

alternate formula: $Var(X) = E[X^2] - (E[X])^2$

standard deviation: $SD(X) = \sqrt{Var(X)}$

Properties

- $Var(cX) = c^2 Var(X)$
- $Var(X + c) = Var(X)$
- $Var(X+Y) \neq Var(X) + Var(Y)$

In Ch. 6, we learn how to expand $Var(X+Y)$

Named distributions (Sections 4.6-4.8)

trial w/prob. of success p

$X \sim \text{Bernoulli}(p)$: $P(X=1) = p$

$[0,1]$ $\qquad P(X=0) = 1-p$

\# of successes in n indep. trials

$X \sim \text{Binom}(n,p)$:

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

$$E[X] = np \qquad Var(X) = np(1-p)$$

$(0,\infty)$

$X \sim \text{Poisson}(\lambda)$:

$$P(X=k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

$$E[X] = \lambda \qquad Var(X) = \lambda$$

Poisson approximation:

$$\text{Binom}(n,p) \approx \text{Poisson}(np)$$

large $\qquad$ small

$\# \text{ of trials until first success}$

$X \sim Geom(p)$: $P(X=k) = (1-p)^{k-1}p$, $k = 1, 2, 3, \ldots$

$p \to (0,1)$

$E[X] = \frac{1}{p}$ $\quad$ $Var(X) = \frac{1-p}{p^2}$

Poisson process (in Section 4.7)

Counting rare events

λ = rate (# of occurrences per unit time)

$N(t)$ = # of occurrences in time interval $[0,t]$

$N = \{N(t) \mid t \geq 0\}$

N is a Poisson process w/ rate λ if

- $N(0) = 0$
- $\underbrace{N(t) - N(s)}_{\#\text{ of occurrences in } (s,t]} \sim \text{Poisson}(\lambda(t-s))$
- N has independent increments

  (occurrences in nonoverlapping time intervals are independent)

4113 Review (no expls)

Chapter 5

Continuous r.v.s (Sections 5.1-5.2, 5.7)

X is continuous if it has a density:

$$P(a \le X \le b) = \int_a^b f_X(x)\,dx$$

nonneg. & integrates to 1

$$P(X \in [x, x+\Delta x]) \approx f_X(x)\,\Delta x$$

$$F_X' = f_X$$

can find f by finding F and differentiating

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

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

If $X \geq 0$: $\quad E[X] = \int_0^\infty P(X > t)\,dt$

If $R(X) = \{0, 1, 2, \ldots\}$: $\quad E[X] = \sum_{n=1}^{\infty} P(X \geq n)$

Named distributions (Sections 5.3–5.6)

Uniform distribution

equally likely to be anywhere in (α, β)

$X \sim \text{Unif}(\alpha, \beta)$:

$\alpha < \beta$

$$f_X(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha < x < \beta, \\ 0 & \text{o.w.} \end{cases}$$

$$E[X] = \frac{\alpha + \beta}{2} \qquad \text{Var}(X) = \frac{(\beta - \alpha)^2}{12}$$

Standard normal distribution

WILL GIVE TABLE OF Φ- VALUES

$Z \sim \text{Norm}(0, 1)$:

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

distribution function: $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} \, du$

$$\Phi(0) = \frac{1}{2}, \qquad \Phi(-x) = 1 - \Phi(x)$$

$$E[Z] = 0 \qquad \text{Var}(Z) = 1$$

Normal distribution

$\sigma \neq 0$

$X \sim \text{Norm}(\mu, \sigma^2)$:

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

$$E[X] = \mu \qquad \text{Var}(X) = \sigma^2$$

$$Z \sim N(0,1) \iff \mu + \sigma Z \sim N(\mu, \sigma^2)$$

Normal approximation

$X \sim \text{Binom}(\underset{\text{large}}{n}, \underset{\text{moderate}}{p})$:

$N(0,1)$

$$X \overset{d}{\approx} \underbrace{\mu}_{E[X]} + \underbrace{\sigma}_{SD(X)} Z$$

Need to use a <u>continuity correction</u>

Exponential distribution

time between consecutive occurrences in a Poisson process w/ rate λ

$\lambda > 0$

$X \sim \mathrm{Exp}(\lambda)$:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0, \\ 0 & \text{o.w.} \end{cases}$$

$$E[X] = \frac{1}{\lambda} \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

$$P(X > t) = e^{-\lambda t} \quad \text{if } t > 0$$

memoryless property: $P(X > t_0 + t \mid X > t_0) = P(X > t)$

WILL GIVE

Gamma function

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

$\alpha > 0$

$\Gamma(1) = 1$, $\Gamma(\alpha+1) = \alpha \Gamma(\alpha)$, $\Gamma(n) = (n-1)!$

Gamma distribution

$X \sim \text{Gamma}(\underbrace{\alpha, \lambda}_{(0,\infty)})$:

$$f(x) = \begin{cases} C \, x^{\alpha-1} e^{-\lambda x} & \text{if } x > 0, \\ 0 & \text{o.w.} \end{cases}$$

where $C = \dfrac{\lambda^\alpha}{\Gamma(\alpha)}$

$E[X] = \dfrac{\alpha}{\lambda}$ $\qquad$ $Var(X) = \dfrac{\alpha}{\lambda^2}$

$\text{Gamma}(1, \lambda) = \text{Exp}(\lambda)$

$\text{Gamma}(n, \lambda) =$ sum of n indep. $\text{Exp}(\lambda)$

covered in Section 6.3.2

Hazard Rates (in Section 5.5)

lifetime of something

$X \geq 0$, density f, dist. func. F

hazard rate function: $\lambda(t) = \dfrac{f(t)}{1 - F(t)}$

$$P(X < t + \Delta t \mid X > t) \approx \lambda(t)\Delta t$$

$$P(X > t) = \exp\left(-\int_0^t \lambda(s)\,ds\right) \quad \text{if} \quad t > 0$$

$$\lambda(t) = \lambda \text{ for all } t \iff X \sim \mathrm{Exp}(\lambda)$$

Chapter 6

Joint distributions (Section 6.1)

Any two r.v.s

joint distribution function:

$$F(x,y) = P(X \le x,\ Y \le y)$$

$$F_X(x) = \lim_{y \to \infty} F(x,y)$$

$$F_Y(y) = \lim_{x \to \infty} F(x,y)$$

Discrete case

Both X, Y discrete.

joint mass function:

$$p(x,y) = P(X = x,\ Y = y)$$

$$p_X(x) = \sum_{y \in R(Y)} p(x,y)$$

$$p_Y(y) = \sum_{x \in R(X)} p(x,y)$$

Continuous case

X, Y jointly continuous.

joint density function

$$P((X,Y)\in D) = \iint_D f(x,y)\,dA$$

$$\frac{\partial^2 F}{\partial x \partial y} = f$$

$$P(X\in[x, x+\Delta x],\ Y\in[y, y+\Delta y]) \approx f(x,y)\,\Delta x \Delta y$$

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

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

Mixed case

X discrete, Y cont.

$$P(X = x_i,\ Y\in A) = \int_A f(x_i, y)\,dy$$

$$p_X(x_i) = \int_{-\infty}^{\infty} f(x_i, y)\,dy$$

$$f_Y(y) = \sum_{x_i\in R(X)} f(x_i, y)$$

Independent r.v.s (Sections 6.2, 6.3)

$X_1, X_2, \ldots$ are <u>independent</u> if

$\{X_1 \in A_1\}, \{X_2 \in A_2\}, \ldots$ are independent

for every choice of $A_1, A_2, \ldots$

Combos of different indep. r.v.s are indep. E.g.,

if $X_1, \ldots, X_7$ are indep, and

$$Y_1 = f(X_1, X_6, X_7)$$
$$Y_2 = g(X_2, X_5)$$

then Y_1 and Y_2 are indep.

All of these are equivalent to independence:

- $F(x,y) = F_X(x) F_Y(y)$ any two r.v.s
- $p(x,y) = p_X(x) p_Y(y)$ discrete case
- $f(x,y) = f_X(x) f_Y(y)$
- $f(x,y) = g(x) h(y)$

continuous case

- $f(x_i, y) = p_X(x_i) f_Y(y)$ mixed case

X, Y cont. and indep:

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

$X \sim \text{Gamma}(\alpha, \lambda)$
$Y \sim \text{Gamma}(\beta, \lambda)$
indep
$\Rightarrow X+Y \sim \text{Gamma}(\alpha+\beta, \lambda)$

$X_1, \dots, X_n \sim \text{Exp}(\lambda)$
indep
$\Rightarrow X_1 + \dots + X_n \sim \text{Gamma}(n, \lambda)$

$X \sim N(\mu_X, \sigma_X^2)$
$Y \sim N(\mu_Y, \sigma_Y^2)$
indep
$\Rightarrow X+Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

$X \sim \text{Binom}(n, p)$
$Y \sim \text{Binom}(m, p)$
indep
$\Rightarrow X+Y \sim \text{Binom}(n+m, p)$

$X \sim \text{Poisson}(\lambda_1)$
$Y \sim \text{Poisson}(\lambda_2)$
indep
$\Rightarrow X+Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$

Covariance and correlation (in Section 6.5)

covariance: $\mathrm{Cov}(X,Y) = E[(X-E[X])(Y-E[Y])]$

alt. formula: $\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$

special case: $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$

correlation: $$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

always in $[-1,1]$

X, Y **uncorrelated** if $\rho(X,Y) = 0$

independent $\Rightarrow$ uncorrelated

Joint normal distribution (in Section 6.5)

X and Y are **jointly normal** if

$aX + bY$ has a normal distribution

whenever a and b are real #'s, not both 0.

- X and Y jointly norm. $\Rightarrow \begin{cases} X \text{ is normal and} \\ Y \text{ is normal} \end{cases}$

- indep. normals are jointly normal

WILL GIVE WHOLE PAGE

X, Y joint normal

joint density:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}$$

alternate form:

$$f(\vec{v}) = \frac{1}{2\pi\sqrt{\det Q}} \exp\left(-\frac{1}{2}(\vec{v}-\vec{m})^T Q^{-1}(\vec{v}-\vec{m})\right)$$

where $\vec{v} = \begin{pmatrix} x \\ y \end{pmatrix}$, $\vec{m} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$, $Q = \begin{pmatrix} \sigma_X^2 & \sigma_X\sigma_Y\rho \\ \sigma_X\sigma_Y\rho & \sigma_Y^2 \end{pmatrix}$

conditional density:

$$f_{X|Y}(x|y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

where $\mu = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)$, $\sigma = \sigma_X(1-\rho^2)$

Conditional distributions (Sections 6.4, 6.5)

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}$$ discrete case

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$ continuous case

$$p_{X|Y}(x_i|y) = \frac{f(x_i,y)}{f_Y(y)}$$

$$f_{Y|X}(y|x_i) = \frac{f(x_i,y)}{p_X(x_i)}$$

mixed case
(X discrete, Y cont.)

4113 Review (no expls)

7.2 Expectations of Sums

$$E[g(X,Y)] = \begin{cases} \sum_{x \in R(X)} \sum_{y \in R(Y)} g(x,y)\,p(x,y) & \text{(both discrete)} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)\,dx\,dy & \text{(jointly continuous)} \\ \sum_{x \in R(X)} \int_{-\infty}^{\infty} g(x,y) f(x,y)\,dy & \text{(mixed)} \end{cases}$$

other related facts:

- $X \le Y \Rightarrow E[X] \le E[Y]$
- $X \le a \Rightarrow E[X] \le a$

7.4 Covariance, Variance of Sums, and Correlation

- X, Y indep $\Rightarrow E[XY] = E[X]E[Y]$
- $\text{Cov}(X,Y) = \text{Cov}(Y,X)$
- $\text{Cov}(X,X) = \text{Var}(X)$

- $\text{Cov}(aX, Y) = a\,\text{Cov}(X, Y)$

- $\text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \text{Cov}(X_i, Y_j)$

- $\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) + \cancel{\sum\sum_{i \neq j} \text{Cov}(X_i, X_j)}$

$$2\sum\sum_{i<j} \text{Cov}(X_i, X_j)$$

Special Case:

$$\text{Var}(X+Y) = \text{Var}(X) + 2\,\text{Cov}(X,Y) + \text{Var}(Y)$$

If $X_1, \ldots, X_n$ are uncorrelated, then

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

7.5 Conditional Expectation

$$E[X \mid Y=y] = \begin{cases} \sum_{x \in R(X)} x\, P(X=x \mid Y=y) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx & \text{if } X \text{ is continuous.} \end{cases}$$

$E[X \mid Y=y]$ is a function of y

$E[X|Y] = g(Y)$, where $g(y) = E[X|Y=y]$

Warning: $g(Y) \neq E[X|Y=Y]$

Have to compute $E[X|Y=y]$ to get an actual function.

$P(A|X=x) = E[1_A | X=x]$

$P(A|X) = E[1_A | X]$

Law of total probability:

$$\boxed{E[X] = E[E[X|Y]]}$$

$E[E[X|Y]] = E[g(Y)]$, where

$g(y) = E[X|Y=y]$

will be a sum if Y is discrete,

integral if Y is continuous

Y discrete:

$$E[X] = \sum_{y \in R(Y)} E[X \mid Y=y]\, P(Y=y)$$

Y continuous:

$$E[X] = \int_{-\infty}^{\infty} E[X \mid Y=y]\, f_Y(y)\, dy$$

<u>Versions for probabilities</u>:

(original law of total probability)

$$P(A) = \sum_{y \in R(Y)} P(A \mid Y=y)\, P(Y=y) \quad \text{if } Y \text{ discrete}$$

$$P(A) = \int_{-\infty}^{\infty} P(A \mid Y=y)\, f_Y(y)\, dy$$

(If X, Y indep, drop the conditioning and turn Y into y.)

$$\boxed{\text{If } X, Y \text{ indep., then } \quad E[f(X,Y) \mid Y=y] = E[f(X,y)]}$$

Special cases:

- X, Y indep $\Rightarrow E[X \mid Y] = E[X]$
- X, Y indep $\Rightarrow P((X,Y) \in D \mid Y=y) = P((X,y) \in D)$

Other helpful facts

- $E[cX|Y] = c\,E[X|Y]$
- $E[X+Y|Z] = E[X|Z] + E[Y|Z]$
- Can pull out known quantities:

$$E[X f(Y)|Y] = f(Y)\,E[X|Y]$$

Special case: $E[f(Y)|Y] = f(Y)$

- Rules for nested conditioning

$$E[E[Z|X,Y]\,|\,X] = E[Z|X]$$

$$E[E[Z|X]\,|\,X,Y] = E[Z|X]$$

- $X \geq 0 \Rightarrow E[X|Y] \geq 0$ (doesn't require independence)
- $E[f(X,Y)|Y=y] = E[f(X,y)|Y=y]$

7.5.4 Conditional Variance

$$\mathrm{Var}(X|Y) = E[(X - E[X|Y])^2 \,|\, Y] \quad \text{(defn)}$$

$$\mathrm{Var}(X|Y) = E[X^2|Y] - (E[X|Y])^2 \quad \text{(alt. formula)}$$

$$\boxed{\mathrm{Var}(X) = E[\mathrm{Var}(X|Y)] + \mathrm{Var}(E[X|Y])}$$

7.7 Moment generating functions

$M_X(t) = E[e^{tX}]$ moment generating func. (MGF)

domain of M_X is all t's such that the expected value exists (it fails to exist when the series or integral is divergent)

Always exists for $t=0$ and $M_X(0) = 1$.

$$\boxed{\underbrace{E[X^n]}_{\text{the "moments" of } X} = M_X^{(n)}(0)}$$

n^{th} derivative of M_X

If X and Y are indep., then

$$M_{X+Y}(t) = M_X(t) M_Y(t)$$

Key thm: If X and Y have the same MGF, then they have the same distribution

$\therefore$ Can use MGF to figure out distribution.
(compute MGF and look it up on the tables)

Table 7.1 Discrete Probability Distribution. WILL GIVE

	Probability mass function, $p(x)$	Moment generating function, $M(t)$	Mean	Variance
Binomial with parameters n, p; $0 \le p \le 1$	$\binom{n}{x} p^x(1-p)^{n-x}$ $x = 0, 1, \ldots, n$	$(pe^t + 1 - p)^n$	np	$np(1-p)$
Poisson with parameter $\lambda > 0$	$e^{-\lambda}\dfrac{\lambda^x}{x!}$ $x = 0, 1, 2, \ldots$	$\exp\{\lambda(e^t - 1)\}$	λ	λ
Geometric with parameter $0 \le p \le 1$	$p(1-p)^{x-1}$ $x = 1, 2, \ldots$	$\dfrac{pe^t}{1-(1-p)e^t}$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$
Negative binomial with parameters r, p; $0 \le p \le 1$	$\binom{n-1}{r-1} p^r(1-p)^{n-r}$ $n = r, r+1, \ldots$	$\left[\dfrac{pe^t}{1-(1-p)e^t}\right]^r$	$\dfrac{r}{p}$	$\dfrac{r(1-p)}{p^2}$

Table 7.2 Continuous Probability Distribution. WILL GIVE

	Probability density function, $f(x)$	Moment generating function, $M(t)$	Mean	Variance
Uniform over (a, b)	$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{e^{tb} - e^{ta}}{t(b-a)}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
Exponential with parameter $\lambda > 0$	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$	$\dfrac{\lambda}{\lambda - t}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
Gamma with parameters $(s, \lambda), \lambda > 0$	$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{s-1}}{\Gamma(s)} & x \ge 0 \\ 0 & x < 0 \end{cases}$	$\left(\dfrac{\lambda}{\lambda - t}\right)^s$	$\dfrac{s}{\lambda}$	$\dfrac{s}{\lambda^2}$
Normal with parameters (μ, σ^2)	$f(x) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x < \infty$	$\exp\left\{\mu t + \dfrac{\sigma^2 t^2}{2}\right\}$	μ	σ^2

8. Limit Theorems

Modes of convergence

- $X_n \to X$ <u>pointwise</u> means:

  $\forall \omega \in \Omega, \ \lim_{n\to\infty} X_n(\omega) = X(\omega)$

- $X_n \to X$ <u>almost surely</u> (or a.s.) means:

  $P(\lim_{n\to\infty} X_n = X) = 1$

- $X_n \to X$ <u>in probability</u> means:

  $\forall \varepsilon > 0, \ \lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$

- $X_n \to X$ <u>in distribution</u> means:

  $\forall x$ at which F_X is continuous,

  $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ whenever F

Thm: Consider the 4 possibilites:

(a) $X_n \to X$ pointwise

(b) $X_n \to X$ a.s.

(c) $X_n \to X$ in probability

(d) $X_n \to X$ in distribution

Then (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d)

Thm If $M_{X_n}(t) \to M_X(t)$, then $X_n \to X$ in distribution.

Prop. 2.1 (Markov's inequality)

If $X \geq 0$ and $a > 0$, then

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Prop. 2.2 (Chebyshev's inequality)

Suppose X has a finite mean. Let $\mu = E[X]$. Then for all $k > 0$,

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

Prop 2.3 If $Var(X) = 0$, then $P(X = E[X]) = 1$.

Thm 2.1 (The weak law of large numbers) (LLN)

Suppose $X_1, X_2, \dots$ are i.i.d. and have a finite mean. Let $\mu = E[X_n]$. Then

$$\frac{X_1 + \cdots + X_n}{n} \to \mu \text{ in probability.}$$

Thm 4.1 (The strong law of large numbers)

Suppose $X_1, X_2, \dots$ are i.i.d. and have a finite mean. Let $\mu = E[X_n]$. Then

$$\frac{X_1 + \cdots + X_n}{n} \to \mu \text{ almost surely}$$

Thm 3.1 (The central limit theorem) (CLT)

Suppose $X_1, X_2, \dots$ are i.i.d., have a finite mean, and a finite positive variance. Let $\mu = E[X_n]$ and $\sigma^2 = \mathrm{Var}(X_n)$. Let $Z \sim N(0,1)$. Then

$$\sqrt{n}\left(\frac{S_n}{n} - \mu\right) \to \sigma Z \text{ in distribution.}$$

Alternate form:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \to Z \text{ in distribution}$$

LLN says $\frac{S_n}{n} \approx \mu \Rightarrow \boxed{S_n \approx n\mu \quad (n \text{ large})}$

CLT says $\frac{S_n - n\mu}{\sqrt{n}\,\sigma} \overset{d}{\approx} Z \Rightarrow \boxed{S_n \overset{d}{\approx} n\mu + \sqrt{n}\,\sigma Z}$

CLT is a higher order approximation.

- Can now use normal approx. on more than binomial.
- Only use cont. correction when approximating discrete r.v.s