STAT 230 - F22
Probability
Full Course Notes

With Prof Audrey Beliveau

These are in-class notes from every lecture; they should have all the content. Hope these help! 😌

Josiah Plett

STAT 230 Notes ☺

① Sept 7th

- Classical Probability: purely statistical
- Relative Frequency: "happened x times last week..."
- Subjective Probability: assumption/guess...

$\bar{A} = A^c = A'$ → "compliment"

$A \cap B = \emptyset$ → "disjoint"

☆ An event in a discrete sample space is any subset of the sample space. ☆ 1 - simple, 2+ - compound

$0 \le P(a_i) \le 1$ and $\sum_{\text{all } i} P(a_i) = 1$

$P(A) = \sum_{a \in A} P(a)$

② Sept 9th (ch 1+2)

- $P(\emptyset) = 0$
- If $A_1, A_2 \cdots$ are disjoint, $P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$ ★
- $P(A^c) = 1 - P(A)$

Odds in... ★

Favour:	Against:
$\frac{P(A)}{1-P(A)}$	$\frac{1-P(A)}{P(A)}$

Finite sample space is "equally likely" if all $P(A_i)$ are equivalent

$|S|$ → size of S

If S is an equally likely sample space: $P(A) = \frac{|A|}{|S|}$

if $A \cap B = \emptyset$: $|A \cup B| = |A| + |B|$

both A and B: $|S(A \cap B)| = |A| \times |B|$

③ Sept 12th Chapter 3.1-3.2

- A permutation (size k) is an ordered subset of k of n objects

Selecting Objects

without replacement: $n^{(k)} = \frac{n!}{(n-k)!}$ ★ $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

with replacement: n^k

④ Sept 14th ch 3.3

Combination ↔ unordered

Permutation ↔ ordered

⑤ Sept 16th ch 3.4,4

De Morgan's Laws

$(A \cup B)^c = A^c \cap B^c$

$(A \cap B)^c = A^c \cup B^c$

n Choose k ★

- $n^{(k)} = \frac{n!}{(n-k)!} = n(n-1)^{(k-1)}$ for $k \ge 1$
- $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n^{(k)}}{k!} = \binom{n}{n-k}$
- $0! = 1$, so $\binom{n}{0} = \binom{n}{n} = 1$
- $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$
- *Binomial theorem:* $(1+x)^n = \binom{n}{0} + \binom{n}{1}x^1 + \ldots + \binom{n}{n}x^n$

$\frac{n!}{n_1!\, n_2! \cdots n_k!}$ ★

Consider n objects of k types. Suppose n_1 objects of type 1; n_k objects of type k. There are $\frac{n!}{n_1! n_2! \cdots n_k!}$ distinguishable arrangements; multinomal coefficient

General Methods ★

① $P(S) = 1$

② $0 \le P(A) \le 1$

③ If $A \subseteq B$, $P(A) \le P(B)$

Series Sums ★

$\sum_{i=0}^{n-1} t^i = \frac{1-t^n}{1-t}$ if $t \ne 1$ | $\sum_{x=0}^{\infty} t^x = \frac{1}{1-t}$ if $|t| < 1$

⑥ Sept 19th ch 4.1-3

Definitions ★

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

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

Independent: $P(A \cap B) = P(A)P(B)$

More than 2? pairwise + all three.

⑦ Sept 21st ch 4.4

$P(A^c | B) = 1 - P(A|B)$

if A_1, A_2 are disjoint:

$P(A_1 \cup A_2 | B) = P(A_1|B) + P(A_2|B)$

Conditional Probability ★

The conditional probability of A given B (when $P(B) > 0$)

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

Independent: $P(A|B) = P(A)$ or $P(B|A) = P(B)$

[if A and B > 0]

⑦ Continued Ch 4.4

Product Rule

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

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

A sequence of sets $A_1, A_2 \cdots A_k$ partition the sample space S if:

$A_i \cap A_j = \emptyset$ for all $i \neq j$, and $\bigcup_{j=1}^{k} A_j = S$

Law of Total Probability ★

Suppose $A_1 \cdots A_k$ partition S. Then for event B:

$P(B) = P(B \cap A_1) + \dots + P(B \cap A_k)$

$P(B) = P(B|A_1)P(A_1) + \dots + P(B|A_k)P(A_k)$

⑧ Sept 23rd Ch 4.5

Bayes Theorem ★

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

- Random variable: a function that assigns a real number to each point in a sample space S. "RV"

- Discrete rv: finite, or countably infinite
- Continuous rv: interval of real numbers

$X \rightarrow$

$X(S) \rightarrow$ • Range: possible values of an rv.

★ • Probability Function (p.f.): $f(x) = P(X = x)$, for all $x \in A$, where X=discrete RV, A= range of X

DISTRIBUTIONS

⑨ Sept 26th ch 5

★ • Cumulative Distribution Function (CDF): $F(x) = P(X \leq x)$, for all $x \in \mathbb{R}$

- X and Y have the same distribution if $F_X(x) = F_Y(x)$ for all $x \in \mathbb{R}$. $X \sim Y$
  - practically, we check over the range

Discrete Uniform ★ $U(a, b)$

if $X \in \{a, a+1, \dots, b-1, b\}$, X has a Discrete Uniform Distribution on a to b, if all values are equally likely.

$U(a, b)$ $\quad f(x) = \frac{1}{b-a+1}$ $\quad F(x) = \begin{cases} 0 & x < a \\ \frac{\lfloor x \rfloor - a + 1}{b - a + 1} & a \leq x \leq b \\ 1 & x > b \end{cases}$

⑩ Sept 28th ch 5

Hypergeometric ★

N objects, where r are "successes" and $N-r$ are "failures." Suppose a subset n is randomly chosen without replacement.

X has hypergeometric distribution if X = # of successes.

$X \sim hyp(N, r, n)$ $\quad f_X(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$, $x = \max(0, n-(N-r)) \leq x \leq \min(r, n)$

⑪ Sept 30th ch 5

- Bernoulli Trial with success probability p is an experiment resulting in success xor failure.

Binomial ★

Perform n Bernoulli Trials with probability p. If X denotes how many successes observed, then X is Binomial (with params n and p)

$X \sim Binomial(n, p)$

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

for $x = 0, 1, \dots, n$

⑫ Oct 3rd ch 5

Negative Binomial ★

Bernoulli trials are repeated with probability p until exactly k successes are seen. Then if X is # of failures before k successes:

$X \sim NB(k, p)$ $\quad f(x) = \binom{x+k-1}{k-1} p^k (1-p)^x$, $x \in \mathbb{N}, x \geq 0$

(12) Oct 3rd Ch 5 (cont'd)

Geometric

If $X \sim NB(1, p)$ then $X \sim Geo(p)$ (winnings only!)

$f(x) = (1-p)^x p, \quad x \in \mathbb{N}$

$F(x) = 1-(1-p)^{\lfloor x \rfloor - 1}$ for $x \geq 0$

(13) Oct 5th Ch 5.7

Poisson

Binomial distribution with very large n and tiny p:

$f(x) = e^{-\mu}\frac{\mu^x}{x!} \approx \binom{n}{x} p^x (1-p)^{n-x}$

where $\mu = np$. $X \sim Poisson(\mu)$

(14) Oct 7th Ch 5.8

Poisson Process: counting the occurrences of an event that happens randomly in space or time.

- Individuality/Singularity
- Independence
- Homogeneity/Uniformity

↳ 3 assumptions → Poisson Process

Poisson Process

If the 3 assumptions [Individuality, Independence, and Homogeneity] hold... let λ denote the rate of event occurrence, and X_t be the number of events seen in time t. Then:

$X_t \sim Poi(\lambda t)$ $\quad f_t(x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$ for $x = 0, 1, \ldots$

"little o" Order Notation

$g(\Delta t)$ is $o(\Delta t)$ as $\Delta t \to 0$ if: $\lim_{\Delta t \to 0} \frac{g(\Delta t)}{\Delta t} = 0$

AKA "$o(\Delta t)$ approaches 0 faster than $g(\Delta t)$"

(15) Oct 17th Ch 7.1, 7.2

Expected Value

Suppose X is a discrete random variable with range A and p.f. $f(x)$. Then $E(X)$ is the "expected value" of X:

$E(X) = \sum_{x \in A} x f(x) = \mu$

AKA
- mean
- first moment

(LOTUS) $E[g(X)] = \sum_{x \in A} g(x) f(x)$ but $g(E(X)) \neq E(g(X))$

Initial cost X... winnings $= y$, net winnings $= y - x$.

(16) Oct 19th ch 7.2-7.3

$E(ag(X)+b) = aE(g(X))+b$

$E(g(X)+h(X)) = E(g(X))+E(h(X))$

$X \sim Binomial(n,p)$ → $E(X) = np$

$X \sim Poi(\mu)$ → $E[X] = \mu$

$X \sim hyp(N, r, n)$ → $E(X) = n\frac{r}{N}$

$X \sim Geo(p)$ → $E(x) = \frac{1-p}{p}$

$X \sim NB(k,p)$ → $E(X) = \frac{k(1-p)}{p}$

Poisson Assumptions

Independence: # of occurrences in non-overlapping intervals are independent.

Individuality: P(2 or more in Δt) approaches 0 faster than P(1 in Δt), as $\Delta t \to 0$.

Uniformity: Probability of one occurrence in $(t, t+\Delta t)$ is $\lambda \Delta t$ for constant λ.

(18) Oct 24th ch 8

- Continuous: X is continuous if its range $X(S)$ is $(a,b) \in \mathbb{R}$.

PDF: Probability Density Function

A crv X has pdf $f(x)$ if:

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$
3. $P(a \leq X \leq b) = \int_a^b f(x)\,dx$

"support" $supp(f) = \{x \in \mathbb{R} : f(x) \neq 0\}$

CDF

crv X has cdf $F(x) = P(X \leq x)$

$F(x) = \int_{-\infty}^{x} f(u)\,du$

$\frac{d}{dx}F(x) = f(x)$

(17) Oct 21st ch 7.4

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

Standard Deviation: $SD(X) = \sqrt{Var(X)}$

- If a and b are constants and $Y = aX + b$, $Var(Y) = a^2 Var(X)$

- Find $F(x)$ from $f(x)$: $F(x) = \int_{-\infty}^{x} f(u)\,du$

p^{th} Quantile

p^{th} quantile of X is the value of $q(p)$ such that $P(X \leq q(p)) = p$ ($100p^{th}$ percentile)

pro-tip: solve $F(x) = p$

(19) Oct 28th 8.2

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

Change of Variables

1. Write the CDF of Y as function of X.
2. Use $F_X(x)$ to find $F_Y(y)$.
3. Find the range of y.

Continuous Uniform

X has continuous uniform distribution on (a, b) if X takes values in (a, b) or [a, b] where all subintervals of a fixed length have the same probability.

$X \sim U(a, b)$

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in (a,b) \\ 0 & \text{otherwise} \end{cases}$$

$$F(x) = \frac{x-a}{b-a}, \quad a \le x \le b$$

(20) Oct 31st 8.2-8.3

$X \sim U(a,b)$

$E(x) = \frac{a+b}{2}$

$Var(x) = \frac{(b-a)^2}{2}$

$X \sim exp(\theta)$

$E(x) = \theta$

$Var(x) = \theta^2$

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

$E(x) = \mu$

$Var(x) = \sigma^2$

C. Exponential

X has exponential distribution if

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & x \ge 0 \\ 0 & x < 0 \end{cases} \qquad X \sim exp(\theta)$$

"scale parameter": $\theta = \frac{1}{\lambda}$ $\quad F(x) = 1 - e^{-x/\theta}$

(21) Nov 2 8.3

Gamma Function: $\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y}\, dy$

[for all $\alpha > 0$]

$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$

if $\alpha \in \mathbb{Z}^+$, $\Gamma(\alpha) = (\alpha-1)!$

(22) Nov 4th 8.4

Theorem Inverse Transform Sampling

Let $F^{-1}(x)$ denote the inverse function of the CDF $F(x)$ of a random variable X.
If $U \sim U(0,1)$, then X defined by $X = F^{-1}(U)$ has CDF $F(x)$.
Then we can simulate "realizations" of X by taking $F^{-1}(\underset{\text{random}}{U})$!

Generalized inverse: $F^{-1}(u) = \inf\{x;\ F(x) \ge u\}$

$\boxed{\text{Infimum}}$ (greatest lower bound)

Nov 7th 8.5

Normal

X has normal (Gaussian) distribution with mean μ and variance σ^2 if the density of X is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)}, \quad x \in \mathbb{R}$$

$X \sim N(\mu, \sigma^2)$ or $X \sim G(\mu, \sigma)$

Properties:

1: Symmetric about the mean.
2: Density is unimodal (1 peak)
3: Mean & Variance are the params.

X is a Standard Normal random variable if $X \sim N(0,1)$. Its frequency is $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$

and CDF: $\Phi(x) = \int_{-\infty}^{x} \varphi(y)\, dy$

STANDARDIZATION:

If $X \sim N(\mu, \sigma^2)$, then defining $Z = \frac{X-\mu}{\sigma}$ gives $Z \sim N(0,1)$

Nov 9th 8.5, 9

Using Standardization: (to find $P(X \le x)$)

1: compute $\frac{x-\mu}{\sigma}$.
2: use standard normal tables to find $P(Z \le \frac{x-\mu}{\sigma})$.
3: this equals $P(X \le x)$!

Joint Probability Function

Suppose $X_1, \ldots, X_n$ are n discrete random variables. Then JPF:

$$f(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

(shorthand for $P(\{X_1 = x_1\} \ldots)$ $x \in X(S) \ldots$)

same sample space

(25) Nov 11th 9.1

$$P((X,Y) \in A) = \sum_{(x,y) \in A} f(x,y)$$

JPF Properties

1. $f(x,y) \ge 0$
2. $\sum_{x,y} f(x,y) = 1$

Marginal Probability Function == pdf

Suppose X and Y are discrete random variables with jpf $f(x,y)$:

$$f_X(x) = P(X = x) = \sum_{y \in Y(S)} f(x,y)$$

Independent iff

$$f(x,y) = f_X(x) f_Y(y) \quad \forall x \in X(S),\ \forall y \in Y(S)$$

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

(26) Nov 14^{th} 9.2, 9.4

Multinomial: more than 2 outcomes.

Multinomial ★

- Trials have k outcomes with probabilities $p_1 + p_2 + \ldots + p_k = 1$.
- Trials repeated n times with $X_1 + X_2 + \ldots + X_k = n$ outcomes.

Then $X_1 \cdots X_k$ have a multinomial distribution.

$$f(x_1, \cdots x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} \quad \text{where } x_1 + \ldots + x_k = n, \; x_i \geq 0$$

Expectation & Variance for Multivariate

$$E(g(X,Y)) = \sum_{(x,y)} g(x,y) f(x,y)$$

if x, y have jpf $f(x,y)$. ★

PROPERTIES:

$$E(a \cdot g_1(X,Y) + b \cdot g_2(X,Y)) = a \cdot E(g_1(X,Y)) + b \cdot E(g_2(X,Y))$$

(27) Nov 16^{th} 9.4

Positive Correlation: $(X - E(X))(Y - E(X)) > 0 \leftrightarrow Cov(X,Y) > 0$

Negative Correlation: $(X - E(X))(Y - E(X)) < 0 \leftrightarrow Cov(X,Y) < 0$

Covariance ★

If X and Y are jointly distributed, then $Cov(X,Y)$ denotes their "Covariance":

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

(28) Nov 18^{th} 9.4, 9.5

★ X and Y are independent $\rightarrow Cov(X,Y) = 0$

↳ converse is not true.

Correlation ★

for LINEAR RELATIONSHIPS ONLY!

$$corr(X,Y) = \rho = \frac{Cov(X,Y)}{SD(X)\,SD(Y)}$$

$-1 \leq \rho \leq 1$

- X and Y are uncorrelated if $Cov(X,Y) = 0$ or $corr(X,Y) = 0$.
- Being uncorrelated doesn't necessarily mean independent.

Linear Combination ★

Suppose $X_1, \ldots X_n$ are jointly distributed RVs with jpf $f(x_1, \ldots, x_n)$. A linear combination of $X_1, \ldots, X_n$ is any RV of the form:

$$\sum_{i=1}^{n} a_i X_i \quad \text{where } a_1, \ldots, a_n \in \mathbb{R}$$

eg

$$\text{Total} = \sum_{i=1}^{n} X_i = T$$

$$\overline{X} = \sum_{i=1}^{n} \frac{1}{n} X_i \quad \leftarrow \text{"sample mean"}$$

$$\overline{X} = \sum_{i=1}^{n} \frac{1}{n} X_i = \frac{1}{n} \sum_{i=1}^{n} X_i = \mu \leftarrow \text{mean of } x_1, \ldots, x_n$$

★ $Cov(X,X) = Var(X)$

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\,Cov(X,Y)$$

if X and Y are independent: $\Rightarrow Var(X+Y) = Var(X-Y) = Var(X) + Var(Y)$

$$Var(\overline{X}) = \frac{\sigma^2}{n}$$

if all r.v.s have the same variance and are independent

(29) Nov 21^{st} 9.6

Variance ★

Let $X_1, \ldots, X_n$ be random variables, and denote $Var(X_i) = \sigma_i^2$, then:

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \sigma_i^2 + 2 \sum_{i=1}^{n} \sum_{j=1+i}^{n} a_i a_j\,Cov(X_i, X_j)$$

If $X_1, \ldots, X_n$ are independent, then $Cov(X_i, X_j) = 0$ $(i \neq j)$ shows:

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

★

If $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$, then:

$$Y \sim N(a\mu + b, a^2\sigma^2)$$

If $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$ independently, then:

$$\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma^2\right)$$

independent, all $N(\mu, \sigma^2)$

$$T = \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

$$\overline{X} \sim N(\mu, \sigma^2/n) \quad ★$$

(30) Nov 23^{rd} 9.7

Indicator RVs ★

Let $A \subset S$ be an event. We say 1_A is the indicator random variable of the event. 1_A is:

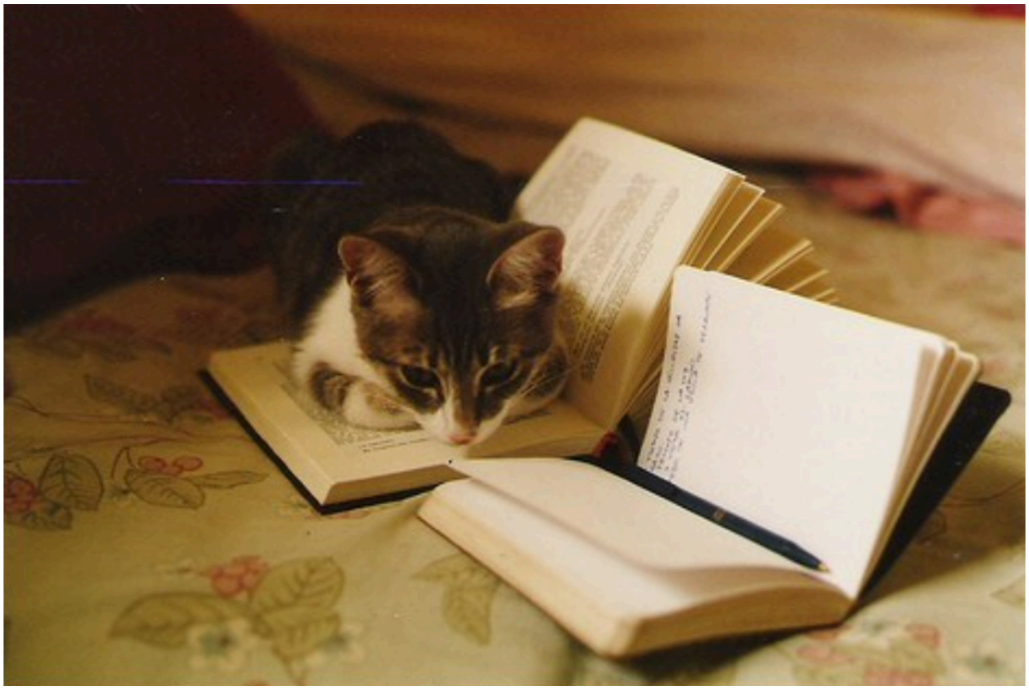
$$1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \in \overline{A} \end{cases}$$

↳ "Bernoulli random variable"

$$E(1_A) = P(A)$$

$$Var(1_A) = P(A)(1 - P(A))$$

Assume bernoulli trial indicator r.v.s are independent



these cats catonify (like personify?) everyone studying stat 230