STATS200 class notes

Erich Trieschman

2021 Fall quarter

1 Combinatorics and probability review

1.1 Counting

1.2 Events and sets

Set operations follow commutative, associative, and distributive laws:

- $E \cup F = F \cup E$

- $E \cap F = F \cap E$ (also written $EF = FE$)

- $(E \cup F) \cup G = E \cup (f \cup G)$

- $(E \cap F) \cap G = E \cap (F \cap G)$

- $(E \cup F) \cap G = (E \cap G) \cup (F \cap G) = E \cap G \cup F \cap G$

- $E \cap F \cup G = (E \cup G) \cap (F \cup G) = E \cup G \cap F \cup G$

DeMorgan's Laws relate the complement of a union to the intersection of complements:

- $(\cup_{i=1}^{n} E_i)^c = \cap_{i=1}^{n} E_i^c$

- $(\cap_{i=1}^{n} E_i)^c = \cup_{i=1}^{n} E_i^c$

1.3 Probability

A **probability space** is defined by a triple of objects $(S, \mathcal{E}, P)$:

- S : Sample space

- $\mathcal{E}$: Set of possible events within the sample space. Set of events are assumed to be θ-field (below)

- P : Probability for each event

A θ-**field** is a collection of subsets $\mathcal{E} \subset S$ that satisfy

- $0 \in \mathcal{E}$

- $E \in \mathcal{E} \Rightarrow E^C \in \mathcal{E}$

- $E_i \in \mathcal{E}$ for $1, 2, \ldots \Rightarrow \cup_{i=1}^{\infty} E_i \in \mathcal{E}$

The **law of total probability** relates marginal probabilities to conditional probabilities. For a partition, $E_1, E_2, \ldots$ of set, S

$$P(A) = \sum_{i=1}^{\infty} P(A \cap E_i)$$

Where a partition implies i) E_i, E_j are pairwise disjoint and ii) $\cup_{i=1}^{\infty} E_i = S$
The **continuity of probability measures** state

$$(i)\ E_1 \subset E_2 \subset \ldots \text{ Let } E_\infty = \cup_i E_i, \text{ then } P(E_n) \longrightarrow P(E_\infty) \text{ as } n \longrightarrow \infty$$
$$(ii)\ E_1 \supset E_2 \supset \ldots \text{ Let } E_\infty = \cap_i E_i, \text{ then } P(E_n) \longrightarrow P(E_\infty) \text{ as } n \longrightarrow \infty$$

1.3.1 Conditional probability

The conditional probability is the probability of one event occurring, given the other event occurring. A reframing of conditional probability (see formula below) is the probability of both events occurring, divided by the marginal probability of one of the events occurring.

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

Bayes Theorem leverages conditional probabilities of measured events to glean conditional probabilities of unmeasured events:

$$P(E_i \mid B) = \frac{P(B \mid E_i)P(E_i)}{\sum_{j=1}^{\infty} P(B \mid E_j)P(E_j)} = \frac{P(B \mid E_i)P(E_i)}{P(B)}$$

Where $E_1, E_2, \ldots$ form a partition of the sample space.

1.3.2 Independence

Events A and B are independent if $P(A \cap B) = P(A)P(B)$

It is possible for events to be pairwise independent, but not mutually independent. For example, toss a pair of dice and let D_1 be the number for die 1 and D_2 be the number for die 2. Define $E_i = \{D_i \leq 2\}$. And define $E_3 = \{3 \leq \max(D_1, D_2) \leq 4\}$. These events are pairwise independent, but $P(E_1 \cap E_2 \cap E_3) = 0$, so they are not mutually independent.

2 Random variables and expected value

Random variables are functions connecting a sample space to real numbers. They are formally defined as

$$\{\omega \in S : X(\omega) \leq t\} \in \mathcal{E}$$

For example, if coin tosses produce a sample space of Heads, Tails, a random variable can be the number of heads.

2.1 Expected value

The expected value (or mean) of a discrete random variable, X, is

$$E(X) = \sum_x x P(X = x)$$

Which can also be written as

$$E(X) = \sum_{x \in S} X(s)p(s), \text{ where } p(s) \text{ is the probability that element } s \in S \text{ occurs:}$$

Proof:

$$E(X) = \sum_i x_i P(X = x_i), \text{ for } E_i = \{X = x_i\} = \{s \in S : X(s) = x_i\}$$

$$= \sum_i x_i \sum_{s \in E_i} p(s) = \sum_i \sum_{s \in E_i} x_i p(s)$$

$$= \sum_i \sum_{s \in E_i} X(s)p(s) = \sum_{s \in S} x_i p(s)$$

This latter equation structure helps build intuition about the linearity of the expected value function and allows us to derive several other properties of expected values. In the general case:

$$E(g(X)) = \sum_i g(x_i) p_X(x_i), \text{ assuming } g(x_i) = y_i$$

Proof:

$$\sum_i g(x_i) p_X(x_i) = \sum_j \sum_{i:g(x_i)=y_j} g(x_i) p_X(x_i) = \sum_j \sum_{i:g(x_i)=y_j} y_j p_X(x_i)$$

$$= \sum_j y_j \sum_{i:g(x_i)=y_j} p_X(x_i) = \sum_j y_j P(g(X) = x_i)$$

$$= E(g(X))$$

And from this general equation we can get two key properties of the expected value:

$(i)\ E(aX + b) = aE(X) + b$

$$E(aX + b) = \sum_{x \in S}(aX(s) + b)p(s) = a\sum_{s \in S} X(s)p(s) + \sum_{s \in S} bp(s) = aE(X) + b$$

$(ii)\ E(X + Y) = E(X) + E(Y)$

$$E(X + Y) = \sum_{s \in S}(X(s) + Y(s))p(s) = \sum_{s \in S} X(s)p(s) + \sum_{s \in S} Y(s)p(s) = E(X) + E(Y)$$

2.2   Variance, covariance, and correlation

The **variance** of X is defined in relation to $E(X) = \mu$ as the expected value of the squared difference between the random variable the mean. The standard deviation, σ is defined as the square root of the variance.

$$Var(X) = E((X - \mu)^2) = \sigma^2$$
$$SD = \sqrt{Var(X)} = \sqrt{\sigma^2} = \sigma$$

Several properties of variance follow from linearity of expectation:

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

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

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

$(ii)\ Var(aX + b) = a^2 Var(X)$

$Var(aX + b) = E((aX + b)^2) - E(aX + b)^2 = E(a^2 X^2 + 2abX + b^2) - (aE(X) + b)^2$

$Var(aX + b) = a^2 E(X^2) + 2abE(X) + b^2 - a^2 E(X)^2 - 2abE(X) - b^2 = a^2 E(X^2) - a^2 E(X)^2 = a^2 (E(X^2) - E(X)^2)$

3   Distribution functions

3.1   Probability mass functions

3.1.1   Binomial distribution

Probability mass function $(Bin(n, p))$: For random variable X, the number of successes in n trials, the probability of observing j successes where each success has probability p is

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

Expected value: Variance:

3.1.2 Geometric distribution

Probability mass function ($Geom(p)$): For random variable X, the number of trials until the first success (included) with probability p is

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

Expected value: Variance:

3.1.3 Negative binomial

Probability mass function ($Geom(p)$): Expected value: Variance:

3.2 Probability density function and cumulative distribution functions

3.2.1 Normal distribution

Probability mass function (): Expected value: Variance:

3.2.2 Exponential distribution

Probability mass function (): Expected value: Variance:

3.2.3 Poisson distribution

Probability mass function (): Expected value: Variance:

3.2.4 Gamma distribution

4 Marginal, joint, and conditional distributions

4.1 Joint distributions

The cumulative density function (cdf) and probability mass function (pmf) satisfy respectively

$$\text{cdf: } F_{X_1,\dots,X_n}(x_1,\dots,x_n) = P(X_i \le x_1,\dots,X_n \le x_n)$$
$$\text{pmf: } f_{X_1,\dots,X_n}(x_1,\dots,x_n) = P(X_1 = x_1,\dots,X_n = x_n)$$

The joint density function f then satisfies, for $E \subset \mathbb{R}^n$,

$$P((X_1,\dots,X_n) \in E) = \int \cdots \int_E f_{X_1,\dots,X_n} dx_1 \dots dx_n$$

When random variables are independent, the joint cdf and pmf satisfy respectively

$$\text{cdf: } P(X_1 \le x_1,\dots,X_n \le x_n) = P(X_1 \le x_1)\dots P(X_n \le x_n)$$
$$\text{pmf: } P(X_1 = x_1,\dots,X_n = x_n) = P(X_1 = x_1)\dots P(X_n = x_n)$$

4.1.1 Distribution of sums of independent random variables

The following combination of marginal distributions is called a **convolution**.
If X and Y have densities, the cdf of $X + Y$ is

$$
\begin{aligned}
F_{X+Y}(t) &= P(X + Y \le t) \\
&= P(X \le t - y) \\
&= \int_{-\infty}^{\infty} P(X \le t - y \mid Y = y) f_x(y) dy, \text{ to get marginal distribution} \\
&= \int_{-\infty}^{\infty} F_x(X \le t - y) f_x(y) dy, \text{ since } X, Y \text{ independent}
\end{aligned}
$$

Likewise, the density of the sum is

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_x(X \leq t - y)f_x(y)dy$$

4.1.2   Expectation of joint distributions

For X, Y joint distribution, $f_{X,Y}(x, y)$, or probability mass function, $p(x, y)$

$$\text{pmf: } E[g(X, Y)] = \sum_s g(X(s), Y(s))p(s)$$

$$= \sum_x \sum_y g(x, y) \sum_{s:X(s)=x,Y(s)=y} p(s)$$

$$= \sum_x \sum_y g(x, y)p(x, y)$$

$$\text{pdf: } E[g(X, Y)] = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} g(x, y)f(x, y)dxdy$$

4.2   Marginal distributions

Marginal density functions or marginal probability mass functions are obtained by integrating or summing out the other variables

$$f_Y(y) = \sum_x yP(Y = y \mid x)$$

4.3   Conditional distributions

Reminder:

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

We can use conditional probabilities to restate the **law of total probability**:

$$P(E) = \int_{-\infty}^{\infty} P(E \mid X = x)f(x)dx$$

5   Moment generating functions

6   Convergence

6.1   Vector products

7   Calculus review

7.1   Infinite sums and series

7.2   Integration

7.3   Derivatives