Lecture 9: Discrete random variables; L&M 3.3 Wed Oct 10, 2018

9.1 Definitions and examples

(i) **Definition:** A random variable X is a real-valued function on the sample space.

(ii) Example: the number of heads in n tosses of a fair coin.

(iii) Example: the number of red-flowered pea-plants out of n, from crosses of $RW \times RW$ pea plants.

(iv) Example: the number of plants grown in (iii) until we see the first white flowered plant.

(v) Example: the number of red fish, in sampling n fish without replacement, from a pond in which there are N fish of which k are red.

(vi) Example: the number of traffic accidents in a large city in a year.

9.2: Discrete Random variables

(i) **Definition:** A random variable X is *discrete* if it can take only at most a countable set of values.

(ii) Example (ii) above: X takes a value in $\mathcal{X} = \{0, 1, 2, ..., n\}$.

(iii) Example (iii) above: ditto.

(iv) X takes values in $\mathcal{X} = \{1, 2, 3, ...\} = \mathcal{Z}_+$.

(v) Example (v) above: X takes values in $\mathcal{X} = \{\max(0, k+n-N), ... \min(k, n)\}$.

(vi) Example (vi) above: X takes values in $\mathcal{X} = \{0, 1, 2, ..\} = \{0\} \cup \mathcal{Z}_+$.

9.3 Probability mass function

(i) **Definition:** The *probability mass function* (p.m.f.) of a discrete random variable X is the set of probabilities $P(X = x)$ for each of the values $x \in \mathcal{X}$ that X can take.

(ii) Example (ii): $P(X = x) = \binom{n}{x}(1/2)^n$ for $x = 0, 1, 2, ...n$.

(iii) Example (iii): $P(X = x) = \binom{n}{x}(3/4)^x(1/4)^{n-x}$ for $x = 0, 1, 2, ...n$.

(iv) Example (iv): $P(X = x) = (3/4)^{(x-1)}.(1/4)$ for $x = 1, 2, 3, ...$:

(v) Example (v): $P(X = x) = \binom{k}{x}\binom{N-k}{n-x}/\binom{N}{n}$ for $x = \max(0, k+n-N), ... \min(k, n)$.

(vi) Example (vi): $P(X = x) = \exp(-\mu)\mu^x/x!$ for $x = 0, 1, 2, 3, 4.....$

9.4 Properties of probability mass functions

(i) $P(X = x) \geq 0$ for each $x \in \mathcal{X}$. (For discrete random variables, in fact $P(X = x) > 0$ for each $x \in \mathcal{X}$.)

(ii) $\sum_x P(X = x) = 1$ where the sum is over all $x \in \mathcal{X}$.

(iii) The subsets E_x of Ω that give $\{X = x\}$ for each x are disjoint events: $P(X \in B) = \sum_{x \in B} P(X = x)$.

9.5 Names of some standard probability mass functions

(i) *Binomial:* Examples (ii) and (iii) are Binomial random variables, $Bin(n, p)$, with index n and *parameter* $p = (1/2)$ and $(3/4)$ respectively.

(ii) *Bernoulli;* If $n = 1$, examples (ii) and (iii) are Bernoulli random variables, $Bernoulli(p)$.

(iii) *Multinomial:* If there are more that two types (for example, number of each of types A, B, AB and O in sample size n from a population) then we have a Multinomial random variable.

(iv) *Geometric:* In example (iv), X is a Geometric random variable.

(iv) *Hypergeometric:* In example (v), X is a Hypergeometric random variable.

(v) *Poisson:* In example (vi), X is a Poisson random variable, with parameter μ, $\mathcal{P}o(\mu)$.

Lecture 10: Continuous random variables: L&M 3.4 Fri Oct 12 (a)

10.1 Definitions and examples

(i) **Definition:** A *continuous* random variable X is one that takes values in $(-\infty, \infty)$. That is, in principle: some values may be impossible.

(ii) Example: A random number between a and b: values in the interval (a, b).

(iii) Example: The waiting time until the bus arrives: values in $(0, \infty)$.

(iv) Example: The USA national debt in 2050: values in $(-\infty, +\infty)$ (in principle).

10.2 The probability density function: definition and basic properties.

(i) **Definition:** The *probability density function* (p.d.f.) of a continuous random variable X is a non-negative function f defined for all values x in $(-\infty, \infty)$ such that for any subset B for which $X \in B$ is an event

$$P(X \in B) \quad = \quad \int_B f(x)\ dx$$

(ii) In fact. (way beyond this course), sets B corresponding to events can be made up of countable unions and intersections of intervals of the form $(a, b]$:

$$P(X \in (a, b]) \quad = \quad P(a < X \leq b) \quad = \quad \int_a^b f(x)\ dx$$

(iii) Note the value at the boundary does not matter:

$$P(X = a) \quad = \quad \int_a^a f(x)\ dx \quad = \quad 0 \qquad \text{for any continuous random variable.}$$

(iv) Note: $f(x) = 0$ is possible for some x-values (see the p.m.f).
For example, if $X \geq 0$ (as in the waiting-time example), $f(x) = 0$, if $x < 0$.

(v) X takes some value in $(-\infty, \infty)$ so

$$1 \quad = \quad P(-\infty < X < \infty) \quad = \quad \int_{-\infty}^{\infty} f(x)\ dx$$

(vi) The *cumulative distribution function* of X is

$$F(x) \quad = \quad P(-\infty < X \leq x) \quad = \quad \int_{-\infty}^{x} f(z)\ dz$$

10.3 Examples: of the probability density function

(i) In general: any non-negative real-valued function f on the real line such that $\int_{-\infty}^{\infty} f(x)\ dx \ = \ 1$.

(ii) Example (ii) above: *Uniform p.d.f*: $U(a, b)$.

$$f(x) \quad = \quad \frac{1}{(b-a)} \text{ for } a \leq x \leq b \text{ and } f(x) \quad = \quad 0 \text{ otherwise.}$$

(iii) Example (iii) above: *Exponential p.d.f* (with rate parameter λ): $\mathcal{E}(\lambda)$.

$$f(x) \quad = \quad \lambda \exp(-\lambda x) \text{ for } x \geq 0 \text{ and } f(x) \quad = \quad 0 \text{ if } x < 0.$$

(iv) Example (iv) above: *Normal p.d.f* (with parameters μ and σ^2): $N(\mu, \sigma^2)$.

$$f(x) \quad = \quad \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \text{ for } -\infty < x < \infty.$$

Lecture 11. Conditioning and functions of a random variable Oct 12 (b)

11.1 Conditioning a discrete random vaiable

Recall $P(X = x)$ is just an event (strictly, defines an event $E_x \subset \Omega$), and $P(X \in B) = \sum_{x \in B} P(X = x)$.

So $P(X \in C \mid X \in B) \;=\; P(X \in B \bigcap C)/P(X \in B)$.

Also $P(X = x \mid X \in B) \;=\; P(X = x)/P(X \in B)$, provided $x \in B$.

11.2 Examples of conditioning a discrete random variable

(i) If X is Poisson, parameter 1: $P(X = x) = e^{-1}1^x/x! \;=\; e^{-1}/x!$, $P(X \geq 2) \;=\; 1 - e^{-1} - e^{-1}$, and $P(X = x \mid X \geq 2) \;=\; (e^{-1}/x!)/(1 - 2e^{-1})$, for $x = 2, 3, 4, ...$

(ii) If X is Bin(11,0.5): $P(X \text{ even}) \;=\; 1/2$ and $P(X = x \mid X \text{ even}) \;=\; 2P(X = x)$ if x is even, 0 otherwise.

(iii) **The forgetting property of the Geometric Distribution**

$P(X = x) \;=\; (1-p)^{x-1}p$ for $x = 1, 2, 3, ...$ Then $P(X > k) \;=\; P(\text{first } k \text{ are failures}) \;=\; (1-p)^k$.

$P(X > k + \ell \mid X > \ell) \;=\; P(X > k + \ell)/P(X > \ell) \;=\; (1-p)^{k+\ell}/(1-p)^{\ell} \;=\; (1-p)^k \;=\; P(X > k)$.

11.3 Conditioning a continuous random varible

Recall $X \in B$ is an event and $P(X \in B) = \int_{x \in B} f_X(x)dx$.

So $P(X \in C | X \in B) \;=\; P(X \in B \cap C)/P(X \in B) \;=\; \int_{B \cap C} f(x)dx / \int_B f(x)dx$

11.4 Examples of conditioning a continuous random variable

(i) Example of a Uniform random variable Suppose X has p.d.f. $f(x) = 1$, $0 \leq x \leq 1$.

$P(X > 0.6 \mid X \leq 0.8) \;=\; P(0.6 < X \leq 0.8)/P(X \leq 0.8) \;=\; 0.2/0.8 \;=\; 0.25$.

(ii) Example for an exponential random variable

Suppose X has p.d.f. $f(x) = 0.5e^{-0.5x}$ on $0 < x < \infty$: $F(x) \;=\; \int^x f(w)dw \;=\; 1 - e^{-0.5x}$.

So $P(X \leq 6 \mid X > 2) = P(2 < X \leq 6)/P(X > 2) \;=\; (e^{-1} - e^{-3})/e^{-1} \;=\; (1 - e^{-2}) \;\approx\; 6/7$.

(iii) The **forgetting property** of the exponential.

Suppose X has p.d.f. $f(x) = \lambda \exp(-\lambda x)$, $\quad 0 < x < \infty$.

Note $F_X(a) = P(X \leq a) \;=\; \int_0^a f(x)\, dx \;=\; (1 - \exp(-\lambda a))$, so $P(X > a) \;=\; \exp(-\lambda a)$. Consider $P(X > a + b \mid X > a) = P(X > a + b)/P(X > a) = \exp(-\lambda(a + b))/\exp(-\lambda a) = \exp(-\lambda b) = P(X > b)$.

11.5 Transformation of a continuous random variable

Example (i): Scaling an exponential random variable.

Suppose $f_X(x) \;=\; \lambda e^{-\lambda x}$ on $x \geq 0$, and let $Y = aX$ $(a > 0)$. What is the pdf of Y?

$$\begin{aligned}
\text{First, } F_X(x) &= \int_0^x \lambda e^{-\lambda w} dw \;=\; [-e^{-\lambda w}]_0^x \;=\; 1 - e^{-\lambda x} \text{ on } x \geq 0. \\
\text{Now, } F_Y(y) &= P(Y \leq y) \;=\; P(aX \leq y) \;=\; P(X \leq y/a) \;=\; F_X(y/a) \;=\; (1 - e^{\lambda y/a}), \\
\text{so } f_Y(y) &= F_Y'(y) \;=\; \frac{d}{dy}(1 - e^{-\lambda y/a}) \;=\; (\lambda/a)e^{-(\lambda/a)y} \text{ on } y \geq 0.
\end{aligned}$$

That is Y is an exponential random variable with parameter λ/a.

Example (ii): Suppose X is Uniform U(0,1). What is the pdf of $Y = X^3$?

$$\begin{aligned}
f_X(x) &= 1,\ 0 \leq x \leq 1; \quad F_X(x) \;=\; x,\ 0 \leq x \leq 1 \\
F_Y(y) \;=\; P(Y \leq y) \;=\; P(X^3 \leq y) \;=\; P(X \leq y^{1/3}) &= F_X(y^{1/3}) \;=\; y^{1/3},\ 0 \leq y \leq 1 \\
f_Y(y) &= \frac{d}{dy}F_Y(y) \;=\; (1/3)y^{-2/3}\ 0 \leq y \leq 1
\end{aligned}$$

Some notes about cumulative distribution functions

1 (i) Definition: For any random variable X, the *cumulative distribution function* is defined as

$$F_X(x) \;=\; P(X \le x) \text{ for } -\infty < x < \infty.$$

(ii) For a discrete random variable with pmf $p_X(x)$, $F_X(b) \;=\; \sum_{x \le b} p_X(x)$.

(iii) For a continuous random variable with pdf $f_X(x)$, $F_X(b) \;=\; \int_{-\infty}^{b} f_X(x)dx$.

(iv) For all random variables, $P(a < X \le b) = F(b) - F(a)$
because $\{X \le b\} \;=\; \{X \le a\} \cup \{a < X \le b\}$ and $\{X \le a\} \cap \{a < X \le b\} \;=\; \Phi$ (empty set).

2 Properties:

(i) F_X is a non-decreasing function: if $a < b$, then $F_X(a) \le F_X(b)$, because $\{X \le a\} \subset \{X \le b\}$.

(ii) $\lim_{b \to \infty} F_X(b) \;=\; 1$, because for any increasing sequence $b_n \to \infty$, $n = 1, 2, 3, ...,$
$\Omega = \{X < \infty\} = \cup\{X \le b_n\}$, so $1 = P(\Omega) \;=\; \lim_{n \to \infty} P(X \le b_n) \;=\; \lim_{n \to \infty} F_X(b_n)$.

(iii) $\lim_{b \to -\infty} F_X(b) \;=\; 0$, because for any decreasing sequence $b_n \to -\infty$, $n = 1, 2, 3, ...,$
$\Phi = \{X = -\infty\} = \cap\{X \le b_n\}$, so $0 = P(\Phi) \;=\; \lim_{n \to \infty} P(X \le b_n) \;=\; \lim_{n \to \infty} F_X(b_n)$.

(iv) F_X is right-continuous. That is, for any b and any decreasing sequence b_n, $n = 1, 2, 3, ...,$ with $b_n \to b$ as $n \to \infty$, $\lim_{n \to \infty} F_X(b_n) = F_X(b)$, because $\{X \le b\} \;=\; \cap\{X \le b_n\}$.

Note $P(X \le b) \;=\; P(X < b) \;+\; P(X = b)$, and $P(X < b) \;=\; \lim_{x \to b^-} F(x)$.

If X is discrete, with $P(X = b) > 0$, F_X will be discontinuous at $x = b$.

3 Case of continuous random variables:

For discrete random variables, $F_X(x)$ is just a set of flat (constant) pieces, with jumps in amount $P(X = x_i)$ at each possible value x_i of X. This is not very useful.

For continuous random variables, the cdf is very useful!

$$F_X(x) \;=\; P(X \le x) \;=\; \int_{-\infty}^{x} f_X(w)dw \quad \text{so} \quad \frac{dF_X(x)}{dx} \;=\; f_X(x).$$

That is, we get the pdf by differentiating the cdf: the cdf is often easier to consider.

Notes and an exercise: Approximating discrete by continuous distributions

Note the geometric and exponential distributions both have the "forgetting" property.

In fact, geometric is just a discrete version of the exponential.

For the *Geom(p)*: $P(X > k) \;=\; (1-p)^k \;\approx\; \exp(-kp)$ for large k and small p.

For the $\mathcal{E}(\lambda)$: $P(X > x) \;=\; \exp(-\lambda x)$

Consider a r.v. T with geometric distribution with $p = 0.02$ and also T with an exponential with rate parameter $\lambda = 0.02$.

(i)Compare $P(T \ge 50)$ under the two models.

(ii) Repeat for $p = 0.001$ (geometric) and $\lambda = 0.001$ (exponential).

(iii) Using either model (p or $\lambda = 0.02$), what is $P(T \ge 100 \mid T \ge 50)$?

(iv) Using either model (p or $\lambda = 0.02$), what is $P(T \le 50 \mid T \le 100)$?

(v) Compare $P(T \le 1)$ for both distributions (p or $\lambda = 0.02$). What about p or $\lambda = 0.2$?