8 Continuous R.V's

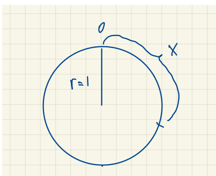
8.1 Probability Distributions in $\Re$

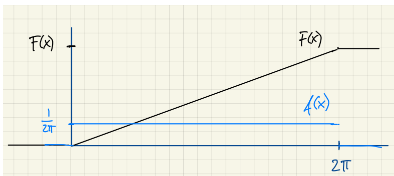
Slide 1

Probability Distributions in $\Re$

§8.1.1 in the book. We skipped §5, 6 and 7 in the book.

Example: consider a roulette wheel of radius[1] 1, and record X = the location where it stops, as distance from 0.





In this case, $\Pr(X = x) = 0$ for any particular value. Instead we work with probabilities of intervals, e.g., $X \le x$.

(Cumulative) Distribution function (cdf): $F(x) = \Pr(X \le x)$.
This is meaningful for *any* r.v.

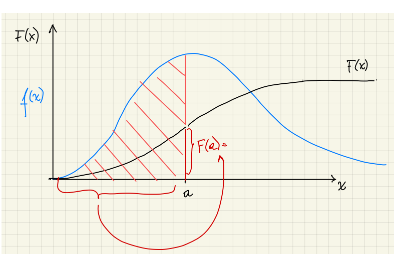
The figure also indicates $f(x) = \frac{d}{dx}F(x)$ – more about $f(x)$ later.

Slide 2

Continuous r.v.: We say X is a *continuous r.v.* if $F(x)$ is a *continuous* function of x.
In that case $\Pr(X < x) = \Pr(X \le x)$, since $\Pr(X = x) = 0$.

(Prob.) Density function (pdf): If there is a function $f(t)$ such

that $F(a) = \int_{-\infty}^{a} f(t)dt$



then $f(t)$ (or $f_X(t)$ if needed for clarity) is called the *p.d.f.* of $F(x)$.
We also use notation like $X \sim f(\cdot)$ or $X \sim F(\cdot)$.

Result: By the fundamental theorem of calculus, $f(x) = \frac{d}{dx}F(x)$.
Probability Mass Function (PMF): The PMF is a concept associated with discrete random variables.
Probability Density Function (PDF): The PDF is a concept associated with continuous random variables.

[1] book uses "circumference" – changed for nicer figure below

8.2 Using the Density Function

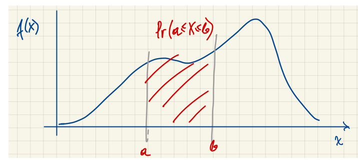
Slide 3

Using the pdf $f(x)$

Using $f(x)$: use $f(x)$ to evaluate $\Pr(X \in A)$ for any event A, e.g. $A = [a, b]$:

$$\Pr(a \le X \le b) = \int_a^b f(x)dx$$

we can think of the pdf as the probability of a little rectangle of width 1, roughly. The pdf, if multiplied by the size of a short interval, gives us the probability of the random.



Interpreting $f(x)$: think of $\Pr(x \le x + \Delta) \approx f(x)\Delta$.
$\Pr(x \le X \le x + \Delta)$

Slide 4

Expectations: recall $E(X) = \sum_x x p_X(x)$ for discrete r.v's.
Similarly, for a continuous r.v.

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

Functions $g(X)$: can show $E\{g(X)\} = \int_{-\infty}^{\infty} g(x)f(x)dx$,
including, e.g., $E(X^k) = \int x^k f(x)dx$, and ...

Variance: if $\mathrm{Var}(X)$ exists, then

$$\begin{aligned} \mathrm{Var}(X) &= E\left\{(X - EX)^2\right\} \\ &= \int_{-\infty}^{\infty} (X - EX)^2 f(x)dx = E(X^2) - \{E(X)\}^2 \end{aligned}$$

Lemma 8.1: If $X \ge 0$, then

$$E(X) = \int_0^{\infty} \Pr(X \ge x)dx = \int_0^{\infty} (1 - F(x))\, dx$$

The proof is analogous to the discrete case ($\to$ Lemma 2.9)

8.3 Examples

Slide 5

Examples

1. Suppose that X is a r.v. with p.d.f.

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2, \\ 0 & \text{otherwise} \end{cases}$$

Find **(a)** C, **(b)** $\Pr(X > 1)$ and, **(c)** $E(X)$.

Solution:

(a) Use $\int_0^2 C(4x - 2x^2)dx = 1$

2x^2 + 2/3x^3

$1/C = \int_0^2 4x - 2x^2 dx = (8 - \frac{2}{3}8) = 8/3 \implies C = 3/8.$

(b) $\Pr(X > 1) = \int_1^2 C(4x - 2x^2)dx = C(2x^2 - \frac{2}{3}x^3)|_1^2 = \frac{1}{2}.$

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

$$= C \int_0^2 x(4x-2x^2)dx = C(\frac{4}{3}x^3 - \frac{1}{2}x^4)|_0^2 = \frac{3}{8}(\frac{32}{3} - \frac{16}{2}).$$

Slide 6 Random Variable X, which represents the outcome of rolling a fair six-sided die. P(X = 1) = 1/6, P(X = 2) = 1/6.

2. For the r.v. X from before, let $Y = X/2$. Find $f_Y(y)$.

Solution: Easiest to start with $F_Y(\cdot)$ and then differentiate.
First, note that $0 < Y < 1$, and

$$\begin{aligned} F_Y(y) = \Pr(Y \le y) &= \Pr(X/2 \le y) = \Pr(X \le 2y) = F_X(2y). \\ \implies f_y(y) &= \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X(2y) \\ &= f_X(2y) \cdot 2 \\ &= C(8y - 8y^2) \cdot 2 = 6(y - y^2), \end{aligned}$$

$0 < y < 1$.

General rule: same works for any monotone

$$Y = g(X) \text{ or } X = g^{-1}(Y).$$

Second last line: $f_Y(y) = f_X[g^{-1}(y)] \cdot \left|\frac{d}{dy}g^{-1}(y)\right|$.

Slide 7

Functions of a R.V.

The same solution works for any r.v. $Y = g(X)$, with an (increasing or decreasing) monotone differentiable function $g(\cdot)$.

Theorem: If $Y = g(X)$, with $g(\cdot)$ strictly monotone and differentiable, then

$$f_Y(y) = f_X[g^{-1}(y)] \left|\frac{d}{dy}g^{-1}(y)\right|.$$

and $f_Y(y) = 0$ if $y \neq g(x)$ for all x.

8.4 Joint Distributions

Slide 8

Joint distributions

When we record multiple numerical summaries of a chance experiment we get multiple jointly distributed r.v's.
We can define *joint* cdf and pdf, similar to before:

pX,Y (a, b) = Pr ((X = a) ∩ (Y = b))
For short, we often write just Pr(X = a, Y = b).

Definition 8.1: The *joint (cumulative) distribution function* (cdf) is

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

We call $f(x, y)$ the *joint (probability) density function* if

F(a) = ∫f(t)dx $$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u, v)\, du dv$$

Similar to the univariate case

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

Finally, if needed for clarity we write $F_{X,Y}(x, y)$ and $f_{X,Y}(x, y)$.

Slide 9

Marginal cdf and pdf: To distinguish the univariate cdf (for X only) from the joint cdf (for X, Y), we include a $_X$ index in

$$F_X(x) = \Pr(X \le x)$$

and refer to it as *marginal distribution*. Same for the *marginal density*, $f_X(x)$.

Marginalization: note The marginal distribution function is a special case of the joint distribution function

$$\begin{aligned} F_X(a) &= \Pr(X \le a) \\ &= \Pr(X \le a, Y \le \infty) = F_{X,Y}(a, \infty) \end{aligned}$$

Since $F_{X,Y}(a, \infty) = \int_{-\infty}^{a} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy\, dx$ we have

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

8.5 Independence

Slide 10

Independence

Definition 8.2: Similar to discrete r.v's we define $X \perp Y$ to be independent if

$$\begin{aligned} F_{X,Y}(x, y) &= \Pr(X \le x, Y \le y) \\ &= \Pr(X \le x) \cdot \Pr(Y \le y) = F_X(x) \cdot F_Y(y) \end{aligned}$$

for all x and y.

Taking derivatives we get an equivalent condition

$$f_{X,Y}(x, y) = f_X(x) f_Y(y).$$

8.6 Examples

Slide 11

Examples

1. Let X and Y be two jointly distributed r.v's with

$$F(x,y) = 1 - e^{-ax} - e^{-by} + e^{-(ax+by)}, \quad x > 0, y > 0$$

Find
(a) $F_X(x)$, **(b)** $F_Y(y)$, **(c)** $f(x,y)$, and **(d)** determine if $X \perp Y$.

Solution: **(a)** $F_X(x) = F(x,\infty) = 1 - e^{-ax}$,

Fx(a) = Pr(X <= a)
= Pr(X <= a, Y <= ∞)
= FX,Y(a, ∞)

(b) $F_Y(y) = F(\infty, y) = 1 - e^{-by}$.

(c) $f(x,y) = ab\, e^{-(ax+by)}$

(d) To show independence note

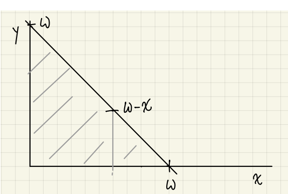
$$\begin{aligned} F_X(x)F_Y(y) &= (1-e^{-ax})(1-e^{-by}) \\ &= 1 - e^{-ax} - e^{-by} + e^{-(ax+by)} \\ &= F(x,y) \end{aligned}$$

$\Longrightarrow X \perp Y$.

Slide 12

#8.2: Let $X \perp Y$ be independent r.v's, $f_X(t) = f_Y(t) = e^{-t}$, $t > 0$.
Find $f_W(w)$ for $W = X + Y$.

Solution: Find $\Pr(W \le w)$.
To determine the following integration limits keep in mind that $f_X(t) = f_Y(t) = 0$ for $t < 0$.



$$\begin{aligned} F_W(w) &= \Pr(X+Y \le w) \\ &= \int_0^\infty \int_{y \le w-x} f_Y(y)dy\, f_X(x)dx \\ &= \int_0^w F_Y(w-x)\, f_X(x)dx. \end{aligned}$$

By differentiation we get

$$\begin{aligned} f_W(w) &= \int_0^w f_Y(w-x)\, f_X(x)dx = \qquad (1) \\ &= \int_0^w e^{-(w-x)-x}dx = w\, e^{-w} \end{aligned}$$

Slide 13

Sum of Exponential Variables

Some comments about this example

- A r.v. with $f_X(x) = \lambda e^{-\lambda x}$, $x > 0$ is known as *exponential* r.v.
  We write $X \sim \text{Exp}(\lambda)$.
  In the example we had $X \sim \text{Exp}(1)$, $Y \sim \text{Exp}(1)$.
- A r.v. with $f_W(w) \propto w^{a-1}\, e^{-bw}$, $w > 0$ is known as a *gamma* r.v., $W \sim \text{Ga}(a,b)$. See more later.
  In the example, $W = X + Y \sim \text{Ga}(2,1)$.
- We have shown: The sum of two independent exponential r.v.'s is a gamma r.v.

Slide 14

Sums of Random Variables

Eq (1) on the previous slide is a general formula for f_{X+Y}, modifying the integration limits in general as implied by the support of f_Y.

Result: Consider two independent r.v.'s X and Y, with p.d.f. $f_X(x)$ and $f_Y(y)$, respectively, and let $S = X + Y$. Then

$$f_S(s) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(s-x)\, dx.$$

Using the result, keep in mind that $f_X(x) = 0$ outside the support of f_X and similarly for f_Y (which typically restricts the integration limits).

Proof: Same as the first few lines of the previous example.

8.7 Conditional Distributions

Slide 15

Conditional Distributions

Recall for two discrete r.v's V, W we defined

$$p_{V|W}(v \mid W = w) = \frac{\Pr(V = v, W = w)}{\Pr(W = w)}$$

(for $\Pr(W = w) > 0$).

The definition does not generalize to continuous r.v's since $\Pr(X = x) = 0$ for a continuous r.v.

But it can also be written as

$$\ldots = \frac{p_{V,W}(v,w)}{p_W(w)}.$$

We use this as a definition!

Conditional density function: for two continuous r.v.'s X, Y,

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

Slide 16

Example: Recall the earlier example with $f(x,y) = abe^{-(ax+by)}$.
We could find

$$f_{X|Y}(x \mid Y = 4) = \frac{abe^{-(ax+4b)}}{be^{-4b}}$$

and, for example

$$\Pr(X \le 3 \mid Y = 4) = \int_{x=0}^{3} \frac{abe^{-(ax+4b)}}{be^{-4b}} = 1 - e^{-3a}.$$

We could have done it faster remembering the independence, $X \perp Y$

$$\implies \Pr(X \le 3 \mid Y = 4) = \Pr(X \le 3) = F_X(3) = 1 - e^{-3a},$$

using $F_X(\cdot)$ from earlier.

Conditional expectation: use $f_{X|Y}$ to define an expectation

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

8.8 Example

Slide 17

Example

Suppose $f(x, y) = \frac{e^{-x/y}e^{-y}}{y}$, $x > 0, y > 0$ for two r.v.'s X and Y.
(a) Find $f_{X|Y}(x \mid y)$; and **(b)** $\Pr(X > 1 \mid Y = y)$.

Solution: **(a)**

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)} = \frac{e^{-x/y-y}/y}{e^{-y}\underbrace{\int_0^{\infty}(1/y)e^{-x/y}dx}_{=1}} = \frac{1}{y}e^{-x/y}$$

(for the integral, see also later, about the exponential distribution).

(b) therefore

$$\Pr(X > 1 \mid Y = y) = \int_1^{\infty} \frac{1}{y}\, e^{-x/y} dx = e^{-1/y}.$$

(again, see also later about the exponential c.d.f).

8.9 The Uniform Distribution

Slide 18

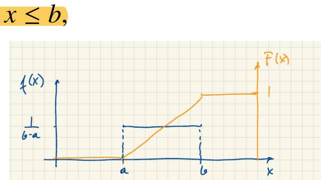
The Uniform Distribution

Assume that X assumes values in an interval $[a, b]$, such that all subintervals of equal length have equal probability.

Then $X \sim \mathrm{Unif}(a, b)$.

Uniform pdf: constant over $a \le x \le b$,

$$f(x) = \begin{cases} 0 & \text{if } x < a \\ 1/(b-a) & \text{if } a \le x \le b \\ 0 & \text{if } x > b, \end{cases}$$



Distribution function: integrating a constant $\to$ straight line:

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } x > b \end{cases}$$

Slide 19

Moments: we find

$$\begin{aligned} E(X) &= \frac{b+a}{2}, \\ E(X^2) &= \int_a^b x\frac{x}{b-a}dx = \frac{b^2+ab+a^2}{3} \end{aligned}$$

$$\implies \mathrm{Var}(X) = E(X^2) - (EX)^2 = \ldots = \frac{(b-a)^2}{12}.$$

8.10 Uniform - Conditional and Order Statistics

Slide 20

Conditional distribution

Lemma 8.2: Assume $X \sim \mathrm{Unif}(a, b)$, then for $x \le d$:

$$\Pr(X \le x \mid X \le d) = \frac{x-a}{d-a}$$

In words, conditional on $X \le d$, X is simply $\mathrm{Unif}(a, d)$.
And similarly, $X \mid X \ge d \sim \mathrm{Unif}(d, b)$.

Slide 21

Order Statistics

Lemma 8.3: Let $X_i \sim \mathrm{Unif}(0, 1)$, $i = 1, \ldots, n$, i.i.d. (independent and identically distributed).

Let $Y_1, \ldots, Y_n$ be the sorted values of $X_1, \ldots, X_n$ (in increasing order) ("order statistic").

Then $E(Y_k) = k/(n+1)$.

Proof: First, consider $Y_1 = \min\{X_1, \ldots, X_n\}$. Then

$$\Pr(Y_1 \ge y) = \Pr(X_1 \ge y, X_2 \ge y, \ldots, X_n \ge y) = (1-y)^n.$$

The last equality is true because of independence of the X_i, and $\Pr(X_i \ge q) = (1 - y)$.

By Lemma 8.1. we have $E(Y_1) = \int_0^1 (1-y)^n dy = \frac{1}{n+1}$.
Extension to $E(Y_k) \to$ Excercise

8.11 Example

Slide 22

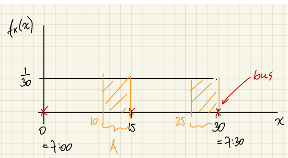
Example

Buses arrive at a specified stop at 15-minute intervals starting at 7am. That is, they arrive at 7, 7:15, 7:30, 7:45, and so on. If a passenger arrives at the stop at a time that is uniformly distributed between 7 and 7:30, find the probability of

(a) $A = \{< 5\text{min wait}\}$; **(b)** $B = \{> 10\text{min wait}\}$.

In the solution, let X = the number of minutes past 7am that the passenger arrives. Then $X \sim \text{Unif}(0, 30)$.

Solution: Note that $f_X(x) = 1/30$, $0 < x < 30$.



$$\begin{aligned} \Pr(A) &= \Pr(\{10 < X < 15\} \cup \{25 < X < 30\}) \\ &= 10/30 = 1/3. \end{aligned}$$

Similarly,

$$\begin{aligned} \Pr(B) &= \Pr(\{0 < X < 5\} \cup \{15 < X < 20\}) \\ &= 10/30 = 1/3. \end{aligned}$$

8.12 The Exponential Distribution

Slide 23

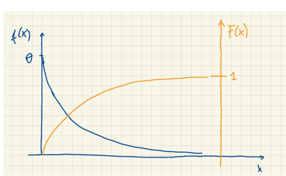
Exponential Distribution

Distribution function: a r.v. $X \geq 0$ follows the exponential distribution if

$$F(x) = \begin{cases} 1 - e^{-\theta x} & \text{for } x \geq 0, \\ 0 & x < 0 \end{cases}$$

Density:

$$f(x) = \frac{d}{dx} F(x) = \theta e^{-\theta x}$$



Moments:

$$\begin{aligned} E(X) &= \int_0^\infty t\,\theta e^{-\theta t} dt = \frac{1}{\theta} \\ E(X^2) &= \int_0^\infty t^2\,\theta e^{-\theta t} dt = \frac{2}{\theta^2} \\ \Longrightarrow \text{Var}(X) &= E\left(X^2\right) - (EX)^2 = \frac{1}{\theta^2} \end{aligned}$$

8.13 Memoryless Property of the Exponential

Slide 24

Memoryless Property

Memoryless property: a remarkable property:

$\Pr(X > s + t \mid X > t) = \Pr(X > s)$

Proof: Will use $\Pr(X > t) = 1 - \Pr(X \leq t) = 1 - F(t) = e^{-\theta t}$.

Note: $1 - F(t)$ is known as *survival function*.

$$\begin{aligned} \Pr(X > s + t \mid X > t) &= \frac{\Pr(X > s + t)}{\Pr(X > t)} \\ &= \frac{e^{-\theta(t+s)}}{e^{-\theta t}} = e^{-\theta s} = \Pr(X > s) \end{aligned}$$

If X is a waiting time (for some event) in min's – after waiting for t minutes, the prob of waiting another s minutes is the same as initially.

Slide 25

Sums of exponential r.v's: Recall #8.2. We showed:

If $X \sim \text{Exp}(1)$ and $Y \sim \text{Exp}(1)$ are independent r.v.'s then $W = X + Y$ is a gamma r.v., $W \sim \text{Ga}(2, 1)$.

In general, if $X_i \sim \text{Exp}(\lambda)$, $i = 1, \ldots, n$, independently, then

$$W = \sum_{i=1}^n X_i \sim \text{Ga}(n, \lambda).$$

Proof is as in #8.2, using induction for $n > 2$.

8.14 Exponential Race

Slide 26

Exponential Race

In words, the minimum of n exponential r.v.'s is exponential with the sum of the rates.

Lemma 8.5: Let $X_i \sim \text{Exp}(\theta_i)$, $i = 1, \ldots, n$, independently. Then

$$U = \min\{X_1, \ldots, X_n\} \sim \text{Exp}(\sum \theta_i)$$

and $\Pr(U = X_i) = \frac{\theta_i}{\sum_{\ell=1}^n \theta_\ell}$.

Proof: for $n = 2$, we get

- for U, using independence of X_1, X_2:

$$\begin{aligned} \Pr(U > x) &= \Pr(X_1 > x, X_2 > x) \\ &= \Pr(X_1 > x) \cdot \Pr(X_2 > x) \\ &= e^{-\theta_1 x} e^{-\theta_2 x} = e^{-(\theta_1 + \theta_2)x} \end{aligned}$$

which is $\Pr(U > x)$ for $U \sim \text{Exp}(\theta_1 + \theta_2)$.

Use induction to show the same for $n > 2$.

Slide 27

- We still have to show the second claim. Note that the joint pdf for (X_1, X_2) is $f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$, and therefore [2]

$$\begin{aligned}\Pr(X_1 < X_2) &= \int_{x_2=0}^{\infty} \int_{x_1=0}^{x_2} \theta_1 e^{-\theta_1 x_1}\, dx_1 \;\; \theta_2 e^{-\theta_2 x_2}\, dx_2 \\ &= \ldots = \frac{\theta_1}{\theta_1 + \theta_2}\end{aligned}$$

Use induction to show the same for $n > 2$.

8.15 Example

Slide 28

Example

Consider a post office that is staffed by two clerks. Suppose that:

- Both clerks are helping a customer and nobody is waiting.
- The two customers, A and B, are served with independent service times, $X_i \sim \text{Exp}(\lambda)$, $i = 1, 2$.
- A third customer, C, enters the post office, and will be served by the first available clerk, again with exponential service time, $X_3 \sim \text{Exp}(\lambda)$.

Find the probability that, of the three customers, C is the last one to leave the post office?

Solution:

- The *remaining* service times Y_i, $i = 1, 2$, for A and B are also $\text{Exp}(\lambda)$ (memoryless property of the Exp).

Slide 29

- When the first customer finishes, two customers are left in the post office. Assume for a moment that A finishes first, leaving then B and C to be served, with (remaining) service times $\text{Exp}(\lambda)$.
- In that case, the probability that C leaves last is therefore $\lambda/(\lambda + \lambda) = 1/2$ (or just argue by symmetry).

  And the same argument if B finishes first

  $\implies$ Pr("C leaves last") = 0.5.

[2] In the lecture the blue & black factors of the integrand were interchanged – not wrong, but possibly misleading