Section 4.3 Continuous random variables

Tuesday, October 11, 2016 1:10 PM

Section 4.3
Continuo...

- **CDF**
  - $F_Y(y)$ is a nondecreasing function.
  - $0 \leq F_Y(y) \leq 1$; this makes sense since $F_Y(y) = P(Y \leq y)$ is a probability!
  - The cdf $F_Y(y)$ in this example takes a "step" at the support points and stays constant otherwise. The height of the step at a particular point is equal to the probability associated with that point. □

CDF PROPERTIES: Let Y be a random variable (discrete or continuous) and suppose that $F_Y(y)$ is the cdf for Y. Then

(i) $F_Y(y)$ satisfies the following:

$$\lim_{y\to-\infty} F_Y(y) = 0 \quad \text{and} \quad \lim_{y\to+\infty} F_Y(y) = 1.$$

(ii) $F_Y(y)$ is a right continuous function; that is, for any real a,

$$\lim_{y\to a^+} F_Y(y) = F_Y(a).$$

(iii) $F_Y(y)$ is a non-decreasing function; that is,

$$y_1 \leq y_2 \Longrightarrow F_Y(y_1) \leq F_Y(y_2).$$

EXERCISE: Graph the cdf for (a) $Y \sim b(5, 0.2)$ and (b) $Y \sim \text{Poisson}(2)$.

4.3 Continuous random variables

TERMINOLOGY: A random variable Y is said to be **continuous** if its cdf $F_Y(y)$ is a continuous function of y.

REMARK: The cdfs associated with discrete random variables are step functions (see Example 4.1). Such functions are not continuous; however, they are still right continuous.

OBSERVATION: We can immediately deduce that if Y is a continuous random variable, then

$$P(Y = y) = 0,$$

for all y. **That is, specific points are assigned zero probability in continuous probability models.** This must be true. If this was not true, and $P(Y = y) = p_0 > 0$, then $F_Y(y)$ would take a step of height p_0 at the point y. This would then imply that $F_Y(y)$ is not a continuous function.

$P(Y<y) = P(Y\le y)$

$P(Y>y) = P(Y\ge y)$

TERMINOLOGY: Let Y be a continuous random variable with cdf $F_Y(y)$. The **probability density function (pdf)** for Y, denoted by $f_Y(y)$, is given by

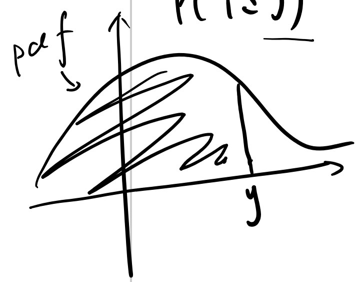
$$f_Y(y) = \frac{d}{dy}F_Y(y),$$

provided that $\frac{d}{dy}F_Y(y) \equiv F'_Y(y)$ exists. Appealing to the Fundamental Theorem of Calculus, we know that

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)dt.$$

$P(Y\le y) = \int_{-\infty}^{y} f_Y(t)dt$

$P(Y\le y)$



These are important facts that describe how the pdf and cdf of a continuous random variable are related. Because $F_Y(y) = P(Y \le y)$, it should be clear that probabilities in continuous models are found by integration (compare this with how probabilities are obtained in discrete models).

PROPERTIES OF CONTINUOUS PDFs: Suppose that Y is a continuous random variable with pdf $f_Y(y)$ and support R. Then

(1) $f_Y(y) > 0$, for all $y \in R$;

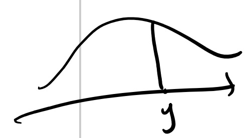
or $f_Y(y) \ge 0$ for $-\infty < y < +\infty$

(2) The function $f_Y(y)$ satisfies

$$\int_R f_Y(y)dy = 1.$$

$\int_{-\infty}^{+\infty} f_Y(y)dy = 1$

$P(Y=y)$
$= P(y\le Y\le y)$
$= \int_y^y f_Y(t)dt$
$= 0$



CONTINUOUS MODELS: Probability density functions serve as **theoretical models** for continuous data (just as probability mass functions serve as models for discrete data). These models can be used to find probabilities associated with future (random) events.

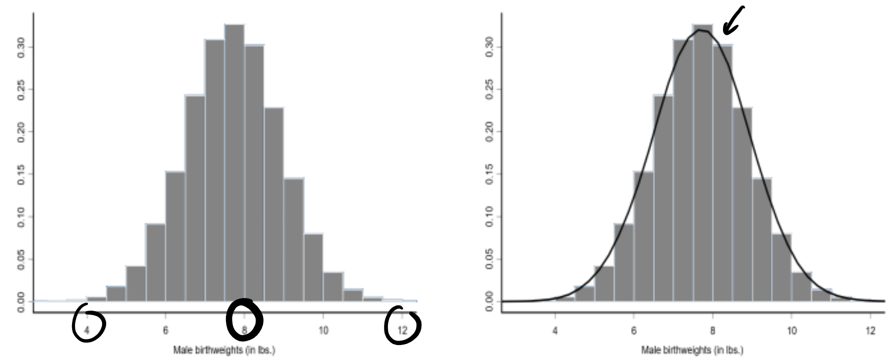


Figure 4.6: *Canadian male birth weight data. The histogram (left) is constructed from a sample of $n = 1250$ subjects. A normal probability density function has been fit to the empirical distribution (right).*

Example 4.2. A team of Montreal researchers who studied the birth weights of five million Canadian babies born between 1981 and 2003 say environmental contaminants may be to blame for a drop in the size of newborn baby boys. A subset ($n = 1250$ subjects) of the birth weights, measured in lbs, is given in Figure 4.6. □

IMPORTANT: Suppose Y is a continuous random variable with pdf $f_Y(y)$ and cdf $F_Y(y)$. The probability of an event $\{Y \in B\}$ is computed by integrating $f_Y(y)$ over B, that is,

$$P(Y \in B) = \int_B f_Y(y)dy,$$

for any $B \subset \mathcal{R}$. If $B = \{y : a \leq y \leq b\}$; i.e., $B = [a, b]$, then

$$\begin{aligned} P(Y \in B) = P(a \leq Y \leq b) &= \int_a^b f_Y(y)dy \\ &= \int_{-\infty}^b f_Y(y)dy - \int_{-\infty}^a f_Y(y)dy \\ &= F_Y(b) - F_Y(a). \end{aligned}$$

$$P(a \leq Y \leq b) = \underbrace{P(Y \leq b)}_{F_Y(b)} - \underbrace{P(Y < a)}_{F_Y(a)}$$

$$= \int_{-\infty}^b f_Y(y)dy - \int_{-\infty}^a f_Y(y)dy$$

Compare these to the analogous results for the discrete case (see page 29 in the notes). In the continuous case, $f_Y(y)$ replaces $p_Y(y)$ and integrals replace sums.

RECALL: We have already discovered that if Y is a continuous random variable, then $P(Y = a) = 0$ for any constant a. This can be also seen by writing

$$P(Y = a) = P(a \leq Y \leq a) = \int_a^a f_Y(y)dy = 0,$$

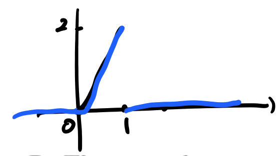
where $f_Y(y)$ is the pdf of Y. An immediate consequence of this is that if Y is continuous,

$$P(a \leq Y \leq b) = P(a \leq Y < b) = P(a < Y \leq b) = P(a < Y < b) = \int_a^b f_Y(y)dy.$$

$= F_Y(b) - F_Y(a)$

Example 4.3. Suppose that Y has the pdf

$$f_Y(y) = \begin{cases} 2y, & 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$



Find the cdf of Y.

Solution. We need to compute $F_Y(y) = P(Y \leq y)$ for all $y \in \mathcal{R}$. There are three cases to consider:

$F_Y(y) = \int_{-\infty}^{y} f_Y(u)\,du$

- when $y \leq 0$,

$$F_Y(y) = \int_{-\infty}^y f_Y(t)dt = \int_{-\infty}^y 0dt = 0;$$

$F_Y(y) = P(Y \leq y) = 0$

positive non-positive

- when $0 < y < 1$,

$$F_Y(y) = \int_{-\infty}^y f_Y(t)dt = \cancel{\int_{-\infty}^0 0dt} + \int_0^y 2tdt = 0 + t^2\Big|_0^y = y^2;$$

$t^2\Big|_0^1$

$P(Y \leq y) = 1$

$(0,1)$ >1

- when $y \geq 1$,

$$F_Y(y) = \int_{-\infty}^y f_Y(t)dt = \cancel{\int_{-\infty}^0 0dt} + \int_0^1 2tdt + \cancel{\int_1^y 0dt} = 0 + 1 + 0 = 1.$$

Putting this all together, we have

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ y^2, & 0 \leq y < 1 \\ 1, & y \geq 1. \end{cases}$$

Find the median $\phi_{0.5}$

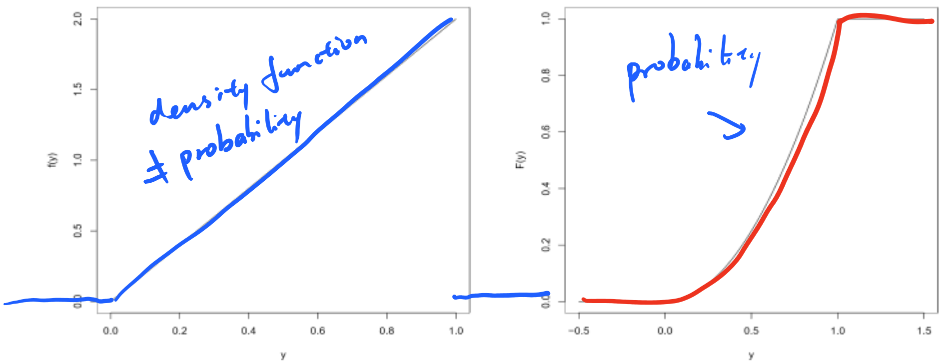
$F_Y(\phi_{0.5}) = 0.5$

$\phi_{0.5}^2 = 0.5$

$\phi_{0.5} = \sqrt{0.5}$

The pdf $f_Y(y)$ and the cdf $F_Y(y)$ are plotted side by side in Figure 4.7.

Exercise: Find (a) $P(0.3 < Y < 0.7)$, (b) $P(Y = 0.3)$, and (c) $P(Y > 0.7)$. □



pdf, $f_Y(y)$ cdf, $F_Y(y)$

Figure 4.7: *Probability density function* $f_Y(y)$ *and cumulative distribution function* $F_Y(y)$ *in Example* 4.3.

Example 4.4. From the onset of infection, the survival time Y (measured in years) of patients with chronic active hepatitis receiving prednisolone is modeled with the pdf

$$f_Y(y) = \begin{cases} \frac{1}{10}e^{-y/10}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

pdf

Find the cdf of Y.

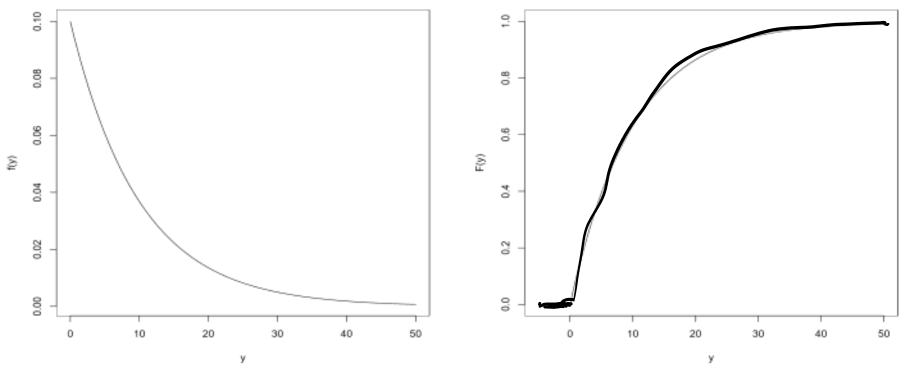
Solution. We need to compute $F_Y(y) = P(Y \leq y)$ for all $y \in \mathcal{R}$. There are two cases to consider:

- when $y \leq 0$,

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)dt = \int_{-\infty}^{y} 0dt = 0;$$

- when $y > 0$,

$$\begin{aligned} F_Y(y) = \int_{-\infty}^{y} f_Y(t)dt &= \int_{-\infty}^{0} 0dt + \int_{0}^{y} \frac{1}{10}e^{-t/10}dt \\ &= 0 + \frac{1}{10}\left(-10e^{-t/10}\right)\Big|_{0}^{y} = 1 - e^{-y/10}. \end{aligned}$$



pdf, $f_Y(y)$ cdf, $F_Y(y)$

Figure 4.8: *Probability density function* $f_Y(y)$ *and cumulative distribution function* $F_Y(y)$ *in Example* 4.4.

Find $\phi_{0.5}$: $F_Y(\phi_{0.5}) = 0.5$

$1 - e^{-\phi_{0.5}/10} = 0.5 \Rightarrow \phi_{0.5} = -10 \times \ln(0.5)$

$\int_{10}^{20} f_Y(t)dt$.

Putting this all together, we have

$$P(Y \le y) = F_Y(y) = \begin{cases} 0, & y \le 0 \\ 1 - e^{-y/10}, & y > 0. \end{cases}$$

The pdf $f_Y(y)$ and the cdf $F_Y(y)$ are plotted side by side in Figure 4.8.

Exercise: What is the probability a patient survives 15 years after being diagnosed? less than 5 years? between 10 and 20 years? □

$P(Y<5) = F_Y(5) = 1 - e^{-5/10}$

$P(10<Y<20) = P(Y<20) - P(Y<10) = F_Y(20) - F_Y(10)$

$P(Y \ge 15) = 1 - P(Y<15)$

$= 1 - F_Y(15)$

$= 1 - (1 - e^{-15/10})$

$= e^{-1.5}$

Example 4.5. Suppose that Y has the pdf

$$f_Y(y) = \begin{cases} cye^{-y/2}, & y \ge 0 \\ 0, & \text{otherwise.} \end{cases}$$

Find the value of c that makes this a valid pdf.

Solution. Because $f_Y(y)$ is a pdf, we know that

$$1 = \int_{-\infty}^{+\infty} f_Y(y)dy = \int_0^\infty f_Y(y)dy = \int_0^\infty cye^{-y/2}dy = 1.$$

$$f_Y(y) = \begin{cases} \frac{1}{4} y e^{-y/2} & y \ge 0 \\ 0 & o.w. \end{cases}$$

Find $\phi_{0.5}$

1. $F_Y(y)$
2. $F_Y(\phi_{0.5}) = 0.5$ solve for $\phi_{0.5}$

solve $\phi_{0.5}$

Using integration by parts with $u = cy$ and $dv = e^{-y/2}dy$, we have

$$1 = \int_0^\infty cye^{-y/2}dy = \underbrace{-2cye^{-y/2}\Big|_0^\infty}_{=0} + \int_0^\infty 2ce^{-y/2}dy$$

$$= 2c(-2)e^{-y/2}\Big|_0^\infty = 0 - (-4c) = 4c.$$

Solving for c, we get $c = 1/4$. □

QUANTILES: Suppose that Y is a continuous random variable with cdf $F_Y(y)$ and let $0 < p < 1$. The **pth quantile** of the distribution of Y, denoted by ϕ_p, solves

$$F_Y(\phi_p) = P(Y \leq \phi_p) = \int_{-\infty}^{\phi_p} f_Y(y)dy = p.$$

The **median** of the distribution of Y is the $p = 0.5$ quantile. That is, the median $\phi_{0.5}$ solves

$$F_Y(\phi_{0.5}) = P(Y \leq \phi_{0.5}) = \int_{-\infty}^{\phi_{0.5}} f_Y(y)dy = 0.5.$$

$F_Y(\phi_{0.5}) = 0.5$
$\shortparallel$
$P(Y \leq \phi_{0.5}) = 0.5$

Another name for the pth quantile is the $100p$**th percentile**.

EXERCISE. Find the median of Y in Examples 4.3, 4.4, and 4.5.

REMARK: For Y discrete, there are some potential problems with the definition that ϕ_p solves $F_Y(\phi_p) = P(Y \leq \phi_p) = p$. The reason is that there may be many values of ϕ_p that satisfy this equation. For example, in Example 4.1, it is easy to see that the median $\phi_{0.5} = 0$ because $F_Y(0) = P(Y \leq 0) = 0.5$. However, $\phi_{0.5} = 0.5$ also satisfies $F_Y(\phi_{0.5}) = 0.5$. By convention, in discrete distributions, the pth quantile ϕ_p is taken to be the smallest value satisfying $F_Y(\phi_p) = P(Y \leq \phi_p) \geq p$.

4.4 Mathematical expectation

4.4.1 Expected value

TERMINOLOGY: Let Y be a continuous random variable with pdf $f_Y(y)$ and support R. The **expected value** of Y is given by

$$E(Y) = \int_R yf_Y(y)dy.$$