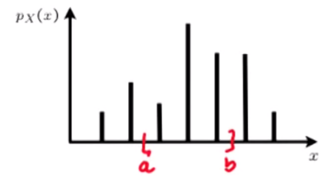
Unit 5 Continuous Random Variables

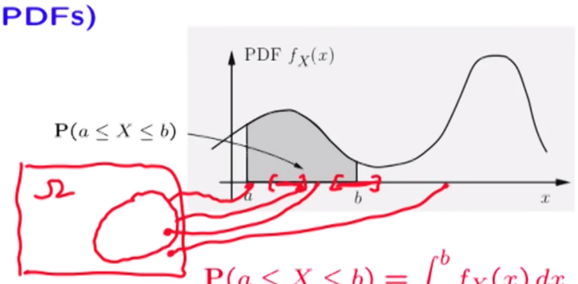
Very much a repetition of unit 4 but with continuous variables instead of discrete variables

Will use more calculus rather than counting/sums

Probability Density Functions (PMFs)
The mass of probability is spread out over a line, calculating the density of this gives us a probability of the interval

Probability density functions (PDFs)





$\mathbf{P}(a \le X \le b) = \sum_{x:\, a \le x \le b} p_X(x)$ $\qquad$ $\mathbf{P}(a \le X \le b) = \int_a^b f_X(x)\, dx$

$p_X(x) \ge 0$ $\qquad$ $\sum_x p_X(x) = 1$ $\qquad$ • $f_X(x) \ge 0$ $\qquad$ • $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$

Definition: A random variable is **continuous if it can be described by a PDF**

$P(1 \le X \le 3 \text{ or } 4 \le X \le 5) = P(1 \le x \le 3) + P(4 \le X \le 5)$

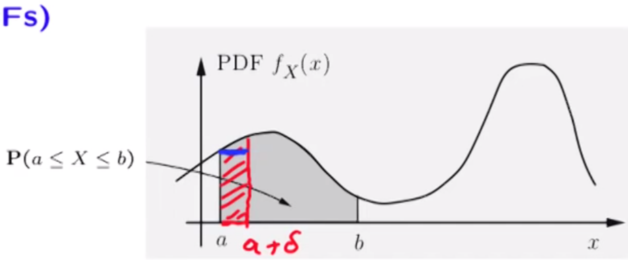
Any particular point has 0 probability but a group of points have a positive probability

Probability density functions (PDFs)

$\delta > 0$, small

$P(a \le X \le a+\delta)$

$\approx f_X(a) \cdot \delta$



$$\boxed{\mathbf{P}(a \le X \le a+\delta) \approx f_X(a) \cdot \delta}$$

$$\boxed{\mathbf{P}(X = a) = 0}$$

$$\mathbf{P}(a \le X \le b) = \int_a^b f_X(x)\,dx$$

$$f_X(x) \ge 0$$

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

$$P(a \le X \le b) = P(X=a) + P(X=b) + P(a < X < b)$$

Exercise: PDFs

4/4 points (graded)

Let X be a continuous random variable with a PDF of the form

$$f_X(x) = \begin{cases} c(1-x), & \text{if } x \in [0,1], \\ 0, & \text{otherwise.} \end{cases}$$

Find the following values.

1. $c =$ 2 ✔
2. $\mathbf{P}(X = 1/2) =$ 0 ✔
3. $\mathbf{P}\big(X \in \{1/k : k \text{ integer}, k \ge 2\}\big) =$ 0 ✔
4. $\mathbf{P}(X \le 1/2) =$ 0.75 ✔

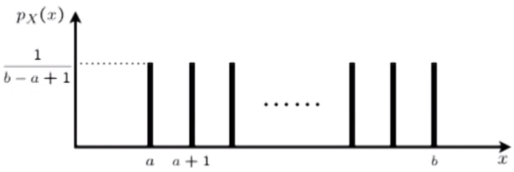
1. We have $1 = \int_{-\infty}^{\infty} f_X(x)\ dx = \int_0^1 c\,(1-x) = c\,(x - x^2/2)\Big|_{0}^{1} = c/2$, and therefore, $c = 2$.

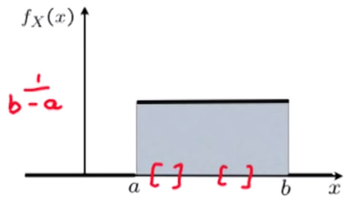
2. Individual points have zero probability.

3. Using countable additivity and the fact that single points have zero probability, we have
$$\mathbf{P}\big(X \in \{1/2, 1/3, 1/4, 1/5, \ldots\}\big) = \sum_{n=2}^{\infty} \mathbf{P}(X = 1/n) = \sum_{n=2}^{\infty} 0 = 0.$$

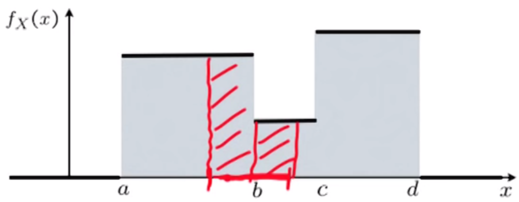
4. $\mathbf{P}(X \leq 1/2) = \int_{-\infty}^{1/2} f_X(x)\ dx = \int_0^{1/2} 2\,(1-x)\ dx = 2\,(x - x^2/2)\Big|_{0}^{1/2} = \frac{3}{4}.$

Example: continuous uniform PDF





- Generalization: piecewise constant PDF



Exercise: Piecewise constant PDF

2/2 points (graded)

Consider a piecewise constant PDF of the form

$$f_X(x) = \begin{cases} 2c, & \text{if } 0 \leq x \leq 1, \\ c, & \text{if } 1 < x \leq 3, \\ 0, & \text{otherwise.} \end{cases}$$

Find the following values.

a) $c =$ 0.25 ✔ **Answer:** 0.25

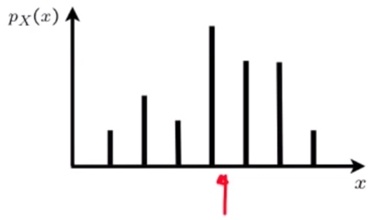
b) $\mathbf{P}(1/2 \leq X \leq 3/2) =$ 0.375 ✔ **Answer:** 0.375

Solution:

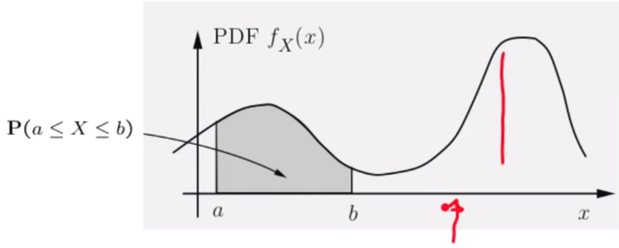
a) The total area under the PDF is the sum of the areas of two rectangles and is equal to $(2c) \cdot 1 + c \cdot 2 = 4c$. Therefore, $c = 1/4$.

b) The total area under the PDF over the interval of interest is the sum of the areas of two smaller rectangles and is equal to $(2c) \cdot (1/2) + c \cdot (1/2) = c \cdot (3/2) = 3/8$.

Expectation/mean of a continuous random variable



$$\mathbf{E}[X] = \sum_x x p_X(x)$$



$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx$$

- **Interpretation:** Average in large number of independent repetitions of the experiment

Fine print:
Assume $\int_{-\infty}^{\infty} |x| f_X(x)\, dx < \infty$

"Centre of gravity" = expectation

Properties of expectations

- If $X \ge 0$, then $\mathbf{E}[X] \ge 0$
- If $a \le X \le b$, then $a \le \mathbf{E}[X] \le b$
- Expected value rule:

$$\mathbf{E}[g(X)] = \sum_x g(x) p_X(x)$$

$$\boxed{\mathbf{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx}$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_x(x)\, dx$$

- Linearity

$$\boxed{\mathbf{E}[aX + b] = a\mathbf{E}[X] + b}$$

Variance and its properties

- **Definition of variance:** $\text{var}(X) = \mathbf{E}[(X-\mu)^2]$

$\mu = E[x]$

- Calculation using the expected value rule, $\mathbf{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$

$$\text{var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f_x(x)\, dx$$

$g(x) = (x-\mu)^2$

Standard deviation: $\sigma_X = \sqrt{\text{var}(X)}$

✓ $\text{var}(aX+b) = a^2\text{var}(X)$

✓ A useful formula: $\text{var}(X) = \mathbf{E}[X^2] - \left(\mathbf{E}[X]\right)^2$

Exercise: Uniform PDF

3/3 points (graded)

Let X be uniform on the interval $[1,3]$. Suppose that $1 < a < b < 3$. Then,

(a) $\mathbf{P}(a \le X \le b) =$ (b-a)*0.5 ✓ **Answer:** (b-a)/2

$(b-a)\cdot 0.5$

(Your answer to part (a) should be an algebraic expression involving a and b.)

(b) $\mathbf{E}[X] =$ 2 ✓ **Answer:** 2

(c) $\mathbf{E}[X^3] =$ 10 ✓ **Answer:** 10

Solution:

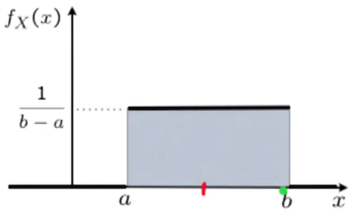
(a) The value of the PDF on the interval $[1,3]$ must be equal to 1/2, so that it integrates to 1. Thus,

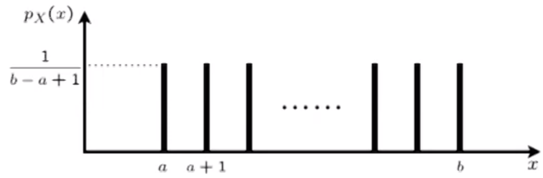
$$\mathbf{P}(a \le X \le b) = \int_a^b \frac{1}{2}\, dx = \frac{b-a}{2}.$$

(b) The expected value of a uniform is the midpoint of its range: $\mathbf{E}[X] = (1+3)/2 = 2$.

(c) Using the expected value rule, $\mathbf{E}[X^3] = \int_1^3 x^3 \cdot \frac{1}{2}\, dx = \frac{1}{2}\cdot\frac{1}{4}x^4\Big|_1^3 = \frac{1}{2}\cdot\frac{1}{4}\cdot(81-1) = 10.$

Continuous uniform random variable; parameters a, b





$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

$$= \int_a^b x \cdot \frac{1}{b-a}\,dx = \frac{a+b}{2}$$

$$\mathbf{E}[X] = \frac{a+b}{2}$$

$$\mathbf{E}[X^2] = \int_a^b x^2 \cdot \frac{1}{b-a}\,dx = \frac{1}{b-a}\left(\frac{b^3}{3} - \frac{a^3}{3}\right)$$

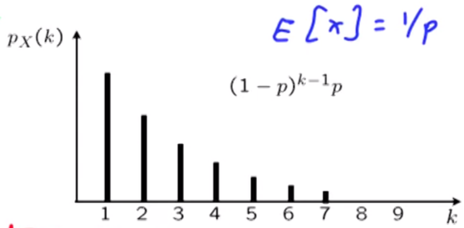
$$\text{var}(X) = \frac{1}{12}(b-a)(b-a+2)$$

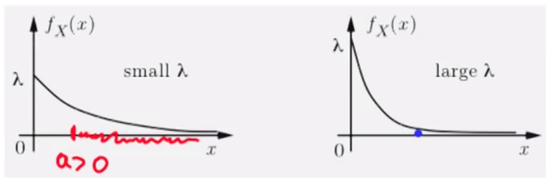
$$\text{var}(X) = E[x^2] - (E[x])^2 = \boxed{(b-a)^2/12} \qquad \sigma = \frac{b-a}{\sqrt{12}}$$

Exponential random variable; parameter $\lambda > 0$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \qquad \int f_X(x)\,dx = 1$$

$E[x] = 1/p$





$$\boxed{P(X \geq a)} = \int_a^{\infty} \lambda e^{-\lambda x}\,dx$$

$$\left[\int e^{ax}\,dx = \frac{1}{a}e^{ax} \quad a \leftrightarrow -\lambda\right]$$

$$= \lambda \cdot \left(-\frac{1}{\lambda}\right) e^{-\lambda x}\Big|_a^{\infty}$$

$$= -e^{-\lambda \cdot \infty} + e^{-\lambda a} = \boxed{e^{-\lambda a}}$$

$$\mathbf{E}[X] = \int_0^{\infty} x \cdot \lambda e^{-\lambda x}\,dx = 1/\lambda$$

$$\mathbf{E}[X^2] = \int_0^{\infty} x^2 \lambda e^{-\lambda x}\,dx = 2/\lambda^2$$

$$\text{var}(X) = E[x^2] - (E[x])^2 = 1/\lambda^2$$

Shows that the probability decreases exponentially beyond a, and if a=0 then the probability equals 1 which is the entire space of probability

When x is zero we have f(x) = lambda since e^0 = 1
The exponential is similar to the discrete geometric and are used in conjunction with one another
The exponential is also used a lot to model the time for an event to happen such as the time for a customer to arrive, a machine to break down, a meteorite to hit a house etc.

Exercise: Exponential PDF

1/2 points (graded)

Let X be an exponential random variable with parameter $\lambda = 2$. Find the values of the following. Use 'e' for the base of the natural logarithm (e.g., enter `e^(-3)` for e^{-3}).

a) $\mathbf{E}[(3X+1)^2] =$ 3.5 ✖ **Answer:** 8.5

b) $\mathbf{P}(1 \le X \le 2) =$ 0.1170 ✔ **Answer:** 0.11702

Solution:

a) By expanding the quadratic, using linearity of expectations, and the facts that $\mathbf{E}[X] = 1/\lambda$ and $\mathbf{E}[X^2] = 2/\lambda^2$, we have

$$\mathbf{E}[(3X+1)^2] = 9\mathbf{E}[X^2] + 6\mathbf{E}[X] + 1 = 9 \cdot \frac{2}{2^2} + 6 \cdot \frac{1}{2} + 1 = \frac{17}{2}.$$

b) We have seen that for $a > 0$, we have $\mathbf{P}(X \ge a) = e^{-\lambda a}$, so that $\mathbf{P}(X \le a) = 1 - e^{-\lambda a}$. Therefore,

$$\mathbf{P}(1 \le X \le 2) = \mathbf{P}(X \le 2) - \mathbf{P}(X \le 1) = (1 - e^{-4}) - (1 - e^{-2}) = e^{-2} - e^{-4}.$$

Will always use capital F to denote a CDF here

Once we know the CDF of a random variable we have enough information to calculate anything we need

Once we have the CDF we can get the PDF (via derivative of CDF) and vice versa (via integrating)

Cumulative distribution function (CDF)

CDF definition: $F_X(x) = \mathrm{P}(X \le x)$

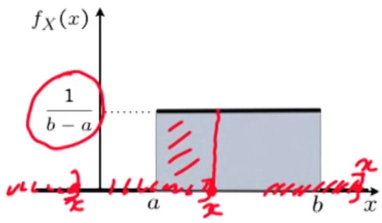
- Continuous random variables:

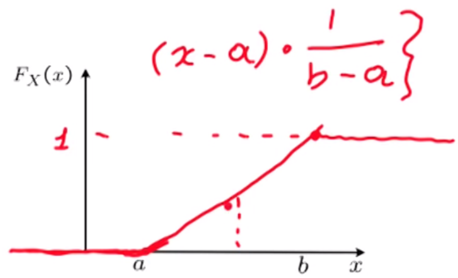
$$F_X(x) = \mathrm{P}(X \le x) = \int_{-\infty}^{x} f_X(t)\, dt$$

3 4

$$\underbrace{P(X \le 4)} = \underbrace{P(X \le 3)} + P(3 < X \le 4)$$

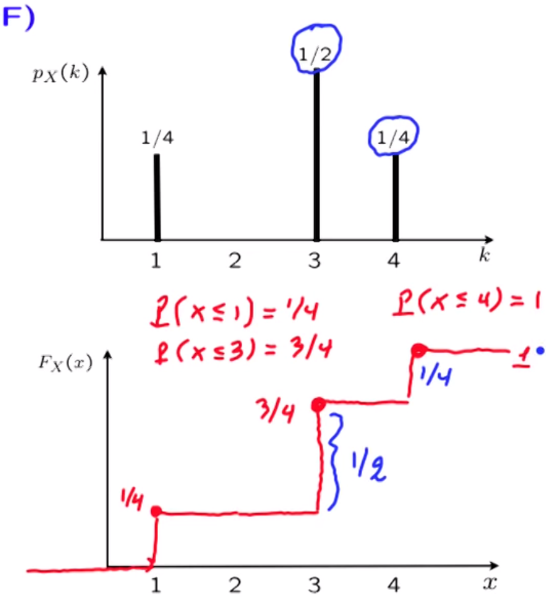
$$\boxed{\frac{dF_X}{dx}(x) = f_X(x)}$$





Cumulative distribution function (CDF)

CDF definition: $F_X(x) = \mathbf{P}(X \le x)$



- Discrete random variables:

$$F_X(x) = \mathbf{P}(X \le x) = \sum_{k \le x} p_X(k)$$

General CDF properties

$$F_X(x) = \mathbf{P}(X \le x)$$

- Non-decreasing   If $y \geq x \Rightarrow F_X(y) \geq F_X(x)$

- $F_X(x)$ tends to 1, as $x \to \infty$

- $F_X(x)$ tends to 0, as $x \to -\infty$

Exercise: Exponential CDF

1/2 points (graded)

Let X be an exponential random variable with parameter 2.

Find the CDF of X. Express your answer in terms of x using standard notation. Use 'e' for the base of the natural logarithm (e.g., enter `e^(-3*x)` for e^{-3x}).

a) For $x \leq 0$, $F_X(x) =$ 0 ✔ **Answer:** 0

0

b) For $x > 0$, $F_X(x) =$ 1 ✖ **Answer:** 1-e^(-2*x)

1

STANDARD NOTATION

Solution:

a) Since X is a nonnegative random variable, $F_X(x) = \mathbf{P}(X \leq x) = 0$ for $x \leq 0$.

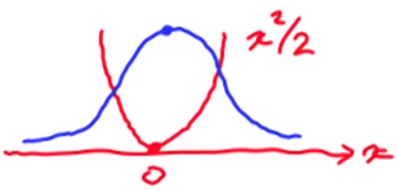
b) We have seen that for an exponential random variable with parameter λ and for any $a > 0$, we have $\mathbf{P}(X \geq a) = e^{-\lambda a}$. Therefore, $F_X(x) = \mathbf{P}(X \leq x) = 1 - \mathbf{P}(X \geq x) = 1 - e^{-\lambda x} = 1 - e^{-2x}$.

Normal (Gaussian) random variables

- Important in the theory of probability
  - Central limit theorem
- Prevalent in applications
  - Convenient analytical properties
  - Model of noise consisting of many, small independent noise terms

Standard normal (Gaussian) random variables

- Standard normal $N(0,1)$: $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$



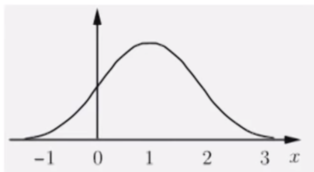
calculus:

$$\int_{-\infty}^{\infty} e^{-x^2/2}\, dx = \sqrt{2\pi}$$

- $\mathbf{E}[X] = 0$
- $\mathrm{var}(X) = 1$ integrate by parts

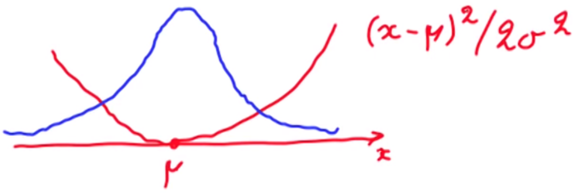
the 1/sqrt(2pi) factor comes from calculus to ensure that the integral equals to 1

General normal (Gaussian) random variables

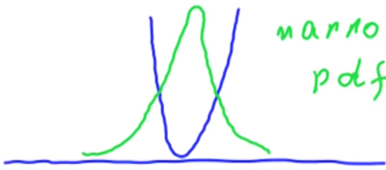


- General normal $N(\mu, \sigma^2)$: $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$

$\sigma > 0$



σ small



narrower pdf

- $\mathbf{E}[X] = \mu$
- $\mathrm{var}(X) = \sigma^2$

Linear functions of a normal random variable

- Let $Y = aX + b \quad X \sim N(\mu, \sigma^2)$

$\mathbf{E}[Y] =$ $a\mu + b$

$\text{Var}(Y) =$ $a^2 \sigma^2$

- Fact (will prove later in this course):

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

- Special case: $a = 0$?

$Y = b$ discrete

$N(b, 0)$

N means Y is still a Normal random variable

Exercise: Normal random variables

1/1 point (graded)

Choose the correct answer below.
According to our conventions, a normal random variable $X \sim N(\mu, \sigma^2)$ is a continuous random variable

- ○ always.
- ◉ if and only if $\sigma \neq 0$.
- ○ if and only if $\mu \neq 0$ and $\sigma \neq 0$.

✔

Solution:

When $\sigma \neq 0$, the distribution of X is described by a PDF, and so X is a continuous random variable. But when $\sigma = 0$, then X has all of its probability assigned to a single point, and therefore it is not a continuous random variable. (For continuous random variables, any single point must have zero probability.)

Calculation of normal probabilities

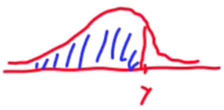
Standard normal tables

- No closed form available for CDF

but have tables, for the standard normal

$Y \sim N(0,1)$

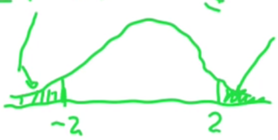
$\Phi(y) = F_Y(y) = P(Y \le y)$



$\Phi(0) = P(Y \le 0) = 0.5$

$\Phi(1.16) = 0.8770$ $\quad \Phi(2.9) = 0.9981$

$\Phi(-2)$



$= 1 - \Phi(2)$

$= 1 - .9772$

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	[illegible]
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	[illegible]
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986

Standardizing a random variable

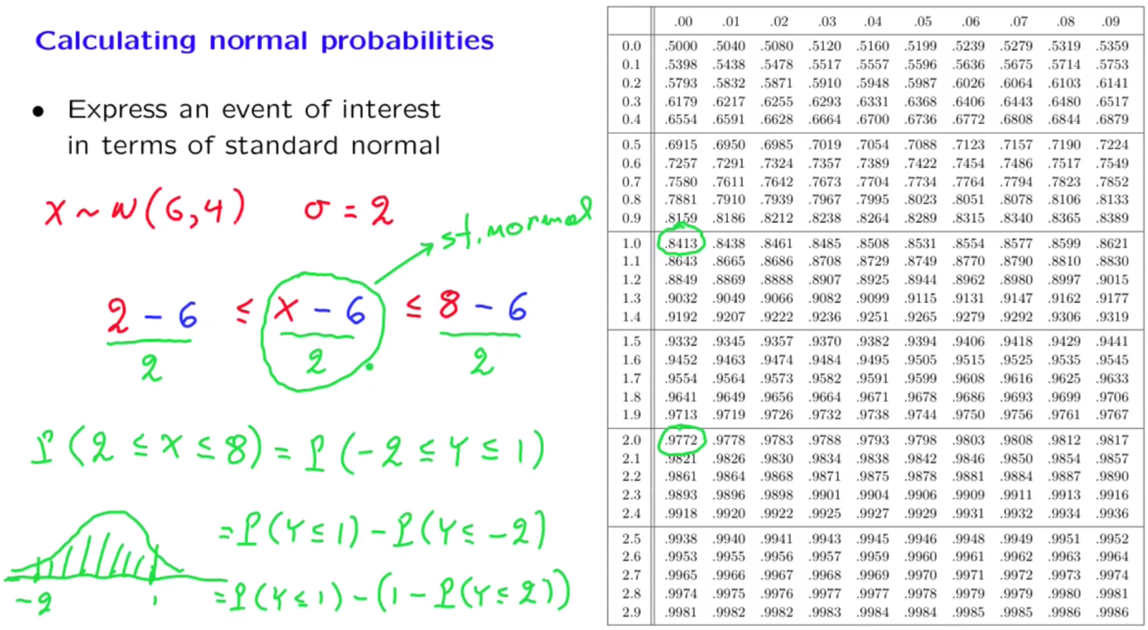
- Let X have mean μ and variance $\sigma^2 > 0$

- Let $Y = \dfrac{X - \mu}{\sigma}$ $\quad E[Y] = 0$ $\quad \mathrm{Var}(Y) = \dfrac{1}{\sigma^2} \mathrm{Var}(X) = 1$

$X = \mu + \sigma Y$

- If also X is normal, then: $Y \sim N(0,1)$

Y measures how many standard deviations it is from the mean



	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986

Standard normal because it is of the form X-mu / sigma
This gives Y so then we can use the standard normal tables

Exercise: Using the normal tables

3/3 points (graded)

Let X be a normal random variable with mean 4 and variance 9.

Use the normal table to find the following probabilities, to an accuracy of 4 decimal places.

Normal Table

Show

a) $\mathbf{P}(X \le 5.2) =$ 0.6554 ✔ **Answer:** 0.6554

b) $\mathbf{P}(X \ge 2.8) =$ 0.6554 ✔ **Answer:** 0.6554

c) $\mathbf{P}(X \le 2.2) =$ 0.2743 ✔ **Answer:** 0.2743

Solution:

a) Note that the standard deviation is 3. Subtracting the mean and dividing by the standard deviation, we obtain
$\mathbf{P}(X \le 5.2) = \mathbf{P}\left(\frac{X-4}{3} \le \frac{5.2-4}{3}\right) = \Phi(0.4) = 0.6554.$

b) Because of the symmetry around the mean, $\mathbf{P}(X \ge 2.8) = \mathbf{P}(X \le 5.2) = 0.6554.$

c) $\mathbf{P}(X \le 2.2) = \mathbf{P}\left(\frac{X-4}{3} \le \frac{2.2-4}{3}\right) = \Phi(-0.6) = 1 - \Phi(0.6) = 1 - 0.7257 = 0.2743.$

Conditioning on an event; Multiple Continuous Random Variables

From discrete to continuous we replace sums with integrals and PMFs with PDFs

Conditional PDF, given an event

$P(A) > 0$

$p_X(x) = \mathbf{P}(X = x)$ $\qquad$ $f_X(x) \cdot \delta \approx \mathbf{P}(x \le X \le x + \delta)$

$p_{X|A}(x) = \mathbf{P}(X = x \mid A)$ $\qquad$ $f_{X|A}(x) \cdot \delta \approx \mathbf{P}(x \le X \le x + \delta \mid A)$

$\mathbf{P}(X \in B) = \sum_{x \in B} p_X(x)$ $\qquad$ $\mathbf{P}(X \in B) = \int_B f_X(x)\, dx$

$\mathbf{P}(X \in B \mid A) = \sum_{x \in B} p_{X|A}(x)$ $\qquad$ $\mathbf{P}(X \in B \mid A) = \int_B f_{X|A}(x)\, dx$ $\qquad$ Def

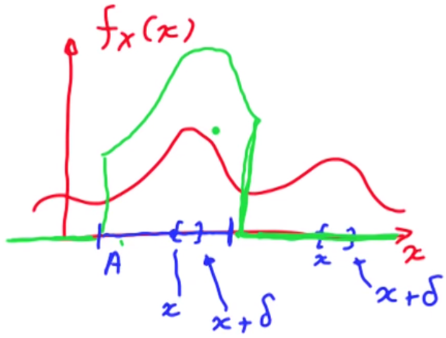
$\sum_x p_{X|A}(x) = 1$ $\qquad$ $\int f_{X|A}(x)\, dx = 1$

Conditional PDF of X, given that $X \in A$

$\mathbf{P}(x \le X \le x + \delta \mid X \in A) \approx f_{X|X \in A}(x) \cdot \cancel{\delta}$

$$= \frac{P(x \le X \le x+\delta,\ X \in A)}{P(A)}$$

$$= \frac{P(x \le X \le x+\delta)}{P(A)} \approx \frac{f_X(x)\,\cancel{\delta}}{P(A)}$$



$$f_{X|X \in A}(x) = \begin{cases} 0, & \text{if } x \notin A \\ \dfrac{f_X(x)}{\mathbf{P}(A)}, & \text{if } x \in A \end{cases}$$

Rescaled so that the total probability under the conditional curve is equal to 1

Conditional expectation of X, given an event

$$\mathbf{E}[X] = \sum_x x p_X(x) \qquad \mathbf{E}[X] = \int x f_X(x)\, dx$$

$$\mathbf{E}[X \mid A] = \sum_x x p_{X|A}(x) \qquad \mathbf{E}[X \mid A] = \int x f_{X|A}(x)\, dx$$

Def

Expected value rule:

$$\mathbf{E}[g(X)] = \sum_x g(x) p_X(x) \qquad \mathbf{E}[g(X)] = \int g(x) f_X(x)\, dx$$

$$\mathbf{E}[g(X) \mid A] = \sum_x g(x) p_{X|A}(x) \qquad \mathbf{E}[g(X) \mid A] = \int g(x) f_{X|A}(x)\, dx$$

Exercise: A conditional PDF

0/1 point (graded)

Suppose that X has a PDF of the form

$$f_X(x) = \begin{cases} 1/x^2, & \text{if } x \geq 1, \\ 0, & \text{otherwise.} \end{cases}$$

For any $x > 2$, the conditional PDF of X, given the event $X > 2$ is

(1/(x^2))/(-1/x) ✖ **Answer:** 2/(x^2)

$$\frac{\frac{1}{x^2}}{-\frac{1}{x}}$$

(Your answer should be an algebraic function of x, in standard notation.)

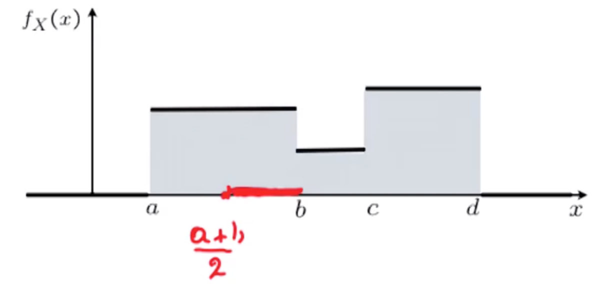
STANDARD NOTATION

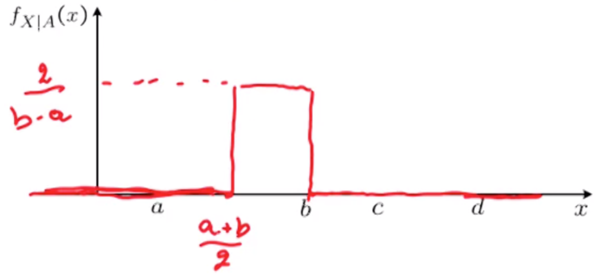
Solution:

The conditional PDF will be a scaled version of the unconditional, of the form $\frac{f_X(x)}{\mathbf{P}(X>2)}$. Now, $\mathbf{P}(X > 2) = \int_2^\infty \frac{1}{x^2}\, dx = -\frac{1}{x}\Big|_\infty^2 = 1/2$, and so the answer is $2/x^2$.

Example

$A: \quad \frac{a+b}{2} \le X \le b$





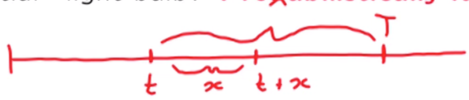
$$\mathbf{E}[X \mid A] = \frac{1}{2} \cdot \frac{a+b}{2} + \frac{1}{2} b$$

$$= \frac{1}{4} a + \frac{3}{4} b$$

$$\mathbf{E}[X^2 \mid A] = \int_{\frac{a+b}{2}}^{b} \frac{2}{b-a} \cdot x^2 \, dx$$

Memorylessness of the exponential PDF

- Do you prefer a used or a new "exponential" light bulb? **Probabilistically identical!**
- Bulb lifetime T: exponential(λ)



$\mathbf{P}(T > x) = e^{-\lambda x}$, for $x \ge 0$

– we are told that $T > t$

– r.v. X: remaining lifetime $= T - t$

$\mathbf{P}(X > x \mid T > t) = e^{-\lambda x}$, for $x \ge 0$

$$= \frac{P(T - t > x, T > t)}{P(T > t)} = \frac{P(T > t + x, T > t)}{P(T > t)} = \frac{P(T > t + x)}{P(T > t)}$$

$$= \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x}$$

This is saying that the probability of a lightbulb lasting x amount of time is the same for a new and a used lightbulb, the memorylessness of the probability ensures this

Electronic items, light bulbs etc., often exhibit the "no ageing" phenomenon. These items do not change properties with usage, but they fail when some external shock like a surge of high voltage, comes along. It can be shown that if these shocks occur according to a Poisson process then the lifetime of the item has exponential distribution.

Practical implications of no ageing :

(a) Since a used component is as good as new (stochastically), there is no advantage in following a policy of planned replacement of used components known to be still functioning.

Memorylessness of the exponential PDF

$f_T(x) = \lambda e^{-\lambda x}$, for $x \geq 0$

$\mathbf{P}(0 \leq T \leq \delta) \approx f_T(0) \cdot \delta = \lambda \delta$

$\mathbf{P}(t \leq T \leq t+\delta \mid T > t) = \approx \lambda \delta$

similar to an independent coin flip,
every δ time steps,
with $\mathbf{P}(\text{success}) \approx \lambda\delta$

The probability of the lightbulb burning out in the next delta tilmestep is always the same
This is the foundation of the Poisson Process

6. Exercise: Memorylessness of the exponential

Bookmark this page

Exercise: Memorylessness of the exponential

2/3 points (graded)

Let X be an exponential random variable with parameter λ.

a) The probability that $X > 5$ is

b) The probabilty that $X > 5$ given that $X > 2$ is

c) Given that $X > 2$, and for a small $\delta > 0$, the probability that $4 \le X \le 4 + 2\delta$ is approximately

Solution:

a) We have seen in the past that for an exponential random variable with parameter λ, $\mathbf{P}(X > a) = e^{-\lambda a}$, and so $\mathbf{P}(X > 5) = e^{-5\lambda}$.

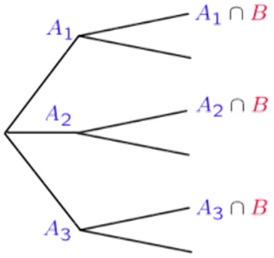
b) Because of the memorylessness property, given that $X > 2$, the remaining time $X - 2$ is again exponential with the same parameter. Thus, $\mathbf{P}(X > 5 \mid X > 2) = \mathbf{P}(X - 2 > 3 \mid X > 2) = \mathbf{P}(X > 3) = e^{-3\lambda}$.

c) By memorylessness, this is the same as the unconditional probability that an exponential takes values in the interval $[2, 2 + 2\delta]$, which is approximately the length, 2δ, of the small interval times the density evaluated at 2, yielding $2\lambda\delta e^{-2\lambda}$.

From lecture, we learned that $P(t < T < (t + \delta) \mid T > t) = \lambda\delta$

In c above the time change is 2delta not delta so it is multiplied by 2.lambda.delta

Total probability and expectation theorems



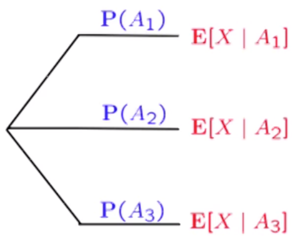
$$\mathrm{P}(B) = \mathrm{P}(A_1)\mathrm{P}(B \mid A_1) + \cdots + \mathrm{P}(A_n)\mathrm{P}(B \mid A_n)$$

$$p_X(x) = \mathrm{P}(A_1)\, p_{X|A_1}(x) + \cdots + \mathrm{P}(A_n)\, p_{X|A_n}(x)$$

$$F_X(x) = P(X \le x) = P(A_1)\, P(X \le x \mid A_1) + \cdots$$
$$= P(A_1)\, F_{X|A_1}(x) + \cdots$$

$$\boxed{f_X(x) = \mathrm{P}(A_1) f_{X|A_1}(x) + \cdots + \mathrm{P}(A_n) f_{X|A_n}(x)}$$

$$\int x f_X(x)\, dx = P(A_1) \int x f_{X|A_1}(x)\, dx + \cdots$$



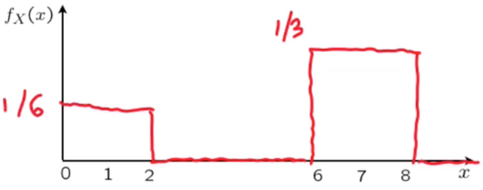
$$\boxed{\mathrm{E}[X] = \mathrm{P}(A_1)\mathrm{E}[X \mid A_1] + \cdots + \mathrm{P}(A_n)\mathrm{E}[X \mid A_n]}$$

Derivative of CDF = PDF
Same formula as discrete case (lower) but now X is a continuous random variable

Example

- Bill goes to the supermarket shortly, with probability 1/3, at a time uniformly distributed between 0 and 2 hours from now; or with probability 2/3, later in the day at a time uniformly distributed between 6 and 8 hours from now

$P(A_1) = \frac{1}{3}$ $f_{X|A_1} \sim unif[0,2]$ $P(A_2) = \frac{2}{3}$ $f_{X|A_2} \sim U[6,8]$



$$f_X(x) = \mathbf{P}(A_1) f_{X|A_1}(x) + \cdots + \mathbf{P}(A_n) f_{X|A_n}(x)$$

$$\mathbf{E}[X] = \mathbf{P}(A_1)\mathbf{E}[X \mid A_1] + \cdots + \mathbf{P}(A_n)\mathbf{E}[X \mid A_n]$$

$$\frac{1}{3} \cdot 1 + \frac{2}{3} \cdot 7$$

Exercise: Total probability theorem II

2/2 points (graded)

On any given day, mail gets delivered by either Alice or Bob. If Alice delivers it, which happens with probability $1/4$, she does so at a time that is uniformly distributed between 9 and 11. If Bob delivers it, which happens with probability $3/4$, he does so at a time that is uniformly distributed between 10 and 12. The PDF of the time X that mail gets delivered satisfies

a) $f_X(9.5) =$ 1/8 ✔ **Answer:** 0.125

b) $f_X(10.5) =$ 4/8 ✔ **Answer:** 0.5

Solution:

The PDF is $1/4$ times a uniform on $[9, 11]$ (of height $1/2$) plus $3/4$ times a uniform on $[10, 12]$ (again of height $1/2$).

a) At time 9.5, only the first uniform is nonzero, yielding $f_X(9.5) = (1/4) \cdot (1/2) = 1/8$.

b) At time 10.5 both uniforms are nonzero, yielding $f_X(10.5) = (1/4) \cdot (1/2) + (3/4) \cdot (1/2) = 1/2$.

Mixed distributions

$$X = \begin{cases} \text{uniform on } [0,2], & \text{with probability } 1/2 \\ 1, & \text{with probability } 1/2 \end{cases}$$

Is X discrete? No

Is X continuous? No

$P(X=1) = 1/2$

X is mixed

Y discrete

Z continuous

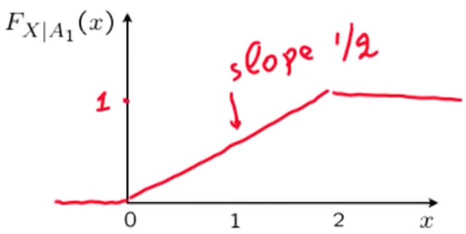
$$X = \begin{cases} Y, & \text{with probability } p \\ Z, & \text{with probability } 1-p \end{cases}$$

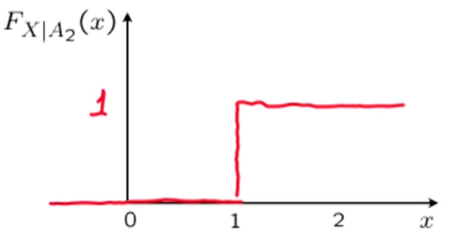
$$F_X(x) = p \cdot P(Y \le x) + (1-p)\, P(Z \le x)$$
$$= p\, F_Y(x) + (1-p)\, F_Z(x)$$

$$E[X] = p\, E[Y] + (1-p)\, E[Z]$$

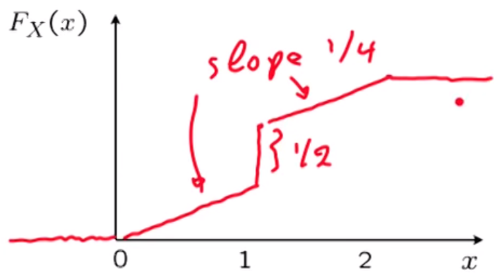
Mixed distributions

$$X = \begin{cases} \text{uniform on } [0,2], & \text{with probability } 1/2 \quad A_1 \\ 1, & \text{with probability } 1/2 \quad A_2 \end{cases}$$





$$F_X(x) = \mathbf{P}(A_1)F_{X|A_1}(x) + \mathbf{P}(A_2)F_{X|A_2}(x)$$



Exercise: A mixed random variable

1/1 point (graded)

A lightbulb is installed. With probability $1/3$, it burns out immediately when it is first installed. With probability $2/3$, it burns out after an amount of time that is uniformly distributed on $[0, 3]$. The expected value of the time until the lightbulb burns out is

1 ✔ **Answer:** 1

Solution:

The expected value of a uniform on $[0, 3]$ is $3/2$. Using the definition of expectation of mixed random variables, the expected value is $\frac{1}{3} \cdot 0 + \frac{2}{3} \cdot \frac{3}{2} = 1.$

Jointly continuous r.v.'s and joint PDFs

$p_X(x)$	$f_X(x)$
$p_{X,Y}(x,y)$	$f_{X,Y}(x,y)$

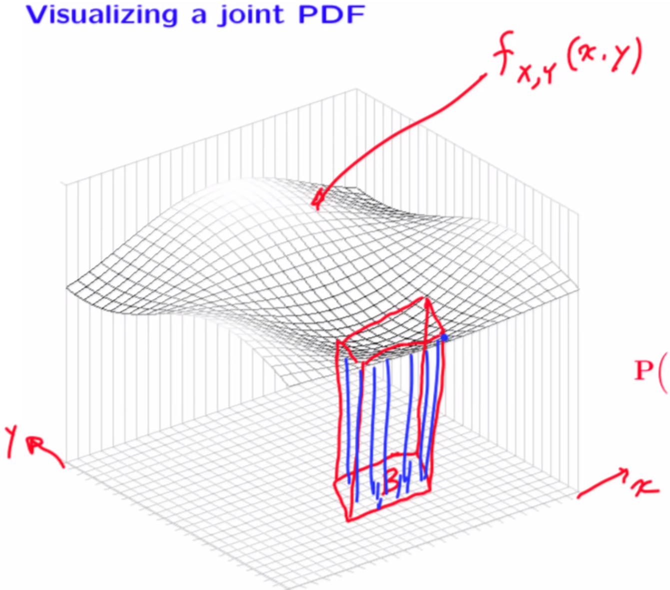
$p_{X,Y}(x,y) = \mathbf{P}(X = x \text{ and } Y = y) \geq 0$ $\qquad$ $f_{X,Y}(x,y) \geq 0$

$$\mathbf{P}\big((X,Y) \in B\big) = \sum\sum_{(x,y)\in B} p_{X,Y}(x,y) \qquad \mathbf{P}\big((X,Y) \in B\big) = \int\int_{(x,y)\in B} f_{X,Y}(x,y)\,dx\,dy$$

$$\sum_x \sum_y p_{X,Y}(x,y) = 1 \qquad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$$

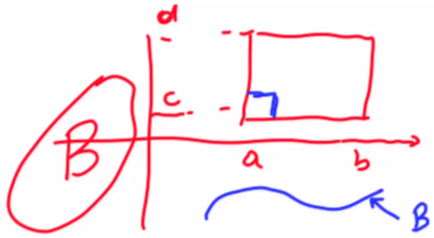
Definition: Two random variables are **jointly continuous** if they can be described by a joint PDF

Visualizing a joint PDF



$$\mathbf{P}\big((X,Y) \in B\big) = \int\int_{(x,y)\in B} f_{X,Y}(x,y)\,dx\,dy$$

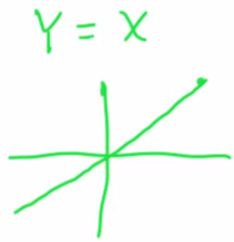
On joint PDFs



$$\mathbf{P}\big((X,Y) \in B\big) = \iint_{(x,y)\in B} f_{X,Y}(x,y)\,dx\,dy$$

$$\mathbf{P}(a \le X \le b,\ c \le Y \le d) = \int_c^d \int_a^b f_{X,Y}(x,y)\,dx\,dy$$

$$\boxed{\mathbf{P}(a \le X \le a+\delta,\ c \le Y \le c+\delta) \approx f_{X,Y}(a,c)\cdot\delta^2}$$



$f_{X,Y}(x,y)$: probability per unit area

$\text{area}(B) = 0 \quad \Rightarrow \quad \mathbf{P}\big((X,Y) \in B\big) = 0$

joint continuity doesn't mean that all r.v.'s are continuous but instead that the probability has to be spread across all their dimensions
Here Y=X is not join continuous because the probability can be defined by one dimension

Exercise: Jointly continuous r.v.'s

2/2 points (graded)

The random variables X and Y are continuous. Is this enough information to determine the value of $\mathbf{P}\big(X^2 = e^{3Y}\big)$?

No ✔ **Answer:** No

The random variables X and Y are jointly continuous. Is this enough information to determine the value of $\mathbf{P}\big(X^2 = e^{3Y}\big)$?

Yes ✔ **Answer:** Yes

Solution:

a) There is no information on the relation between the two random variables. If, for example, $X = \sqrt{e^{3Y}}$, the probability is 1, whereas if $X = \sqrt{e^{3Y}} + 1$, then the probability is zero.

b) The set of points on the x-y plane that correspond to the event $X^2 = e^{3Y}$ is a one-dimensional curve, which has zero area, and therefore zero probability.

Exercise: From joint PDFs to probabilities

8/8 points (graded)

a) The probability of the event that $0 \le Y \le X \le 1$ is of the form $\int_a^b \left(\int_c^d f_{X,Y}(x, y)\ dx \right)\ dy$.

Find the values of a, b, c, d. Each one of your answers should be one of the following: 0, x, y, or 1.

$a =$ 0 ✔ **Answer:** 0

0

$b =$ 1 ✔ **Answer:** 1

1

$c =$ y ✔ **Answer:** y

y

$d =$ 1 ✔ **Answer:** 1

1

b) The probability of the event that $0 \le Y \le X \le 1$ is also of the form $\int_a^b \left(\int_c^d f_{X,Y}(x, y)\ dy \right)\ dx$. Note the different order of integration as compared to part (a).

Find the values of a, b, c, d. Each one of your answers should be one of the following: 0, x, y, or 1.

$a =$ 0 ✔ **Answer:** 0

0

$b =$ 1 ✔ **Answer:** 1

1

$c =$ 0 ✔ **Answer:** 0

0

$d =$ x ✔ **Answer:** x

x

Solution:

a) For any given $y \in [0, 1]$, x ranges from y to 1, yielding $\int_0^1 \int_y^1 f_{X,Y}(x, y)\ dx\, dy$.

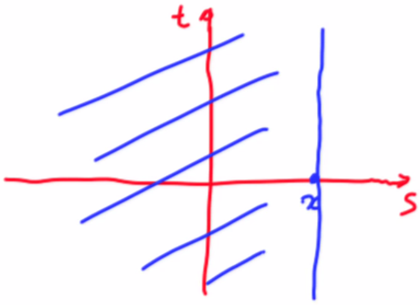
b) For any given $x \in [0, 1]$, y ranges from 0 to x, yielding $\int_0^1 \int_0^x f_{X,Y}(x, y)\ dy\, dx$.

From the joint to the marginals

$$p_X(x) = \sum_y p_{X,Y}(x,y) \qquad f_X(x) = \int f_{X,Y}(x,y)\,dy$$

$$p_Y(y) = \sum_x p_{X,Y}(x,y) \qquad f_Y(y) = \int f_{X,Y}(x,y)\,dx$$

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} \underbrace{\left[\int_{-\infty}^{\infty} f_{X,Y}(s,t)\,dt\right]}\,ds$$

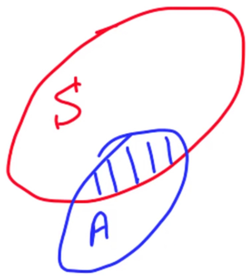


$$f_X(x) = \frac{dF_X}{dx}(x) = [\ \ ]$$

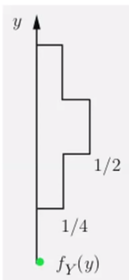
Uniform joint PDF on a set S

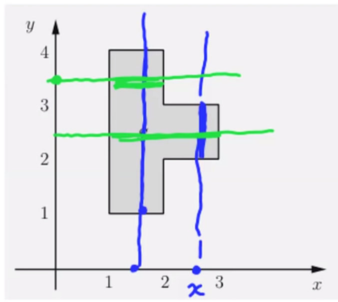
$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{\text{area of } S}, & \text{if } (x,y) \in S, \\ 0, & \text{otherwise.} \end{cases}$$

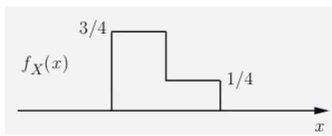
$$f_{X,Y} = \frac{1}{4}$$



$$P(A) = \frac{\text{area}(A \cap S)}{\text{area}(S)}$$







Exercise: Finding a marginal PDF

1/1 point (graded)

The random variables X and Y are described by a uniform joint PDF of the form $f_{X,Y}(x, y) = 3$ on the set $\{(x, y) \mid 0 \le x \le 1,\ 0 \le y \le 1,\ y \le x^2\}$.

Then, $f_X(0.5) =$ 0.75 ✔ **Answer:** 0.75

Solution:

For any $x \in [0, 1]$, and using also the fact that the PDF is zero outside the specified set of x-y pairs, we have $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dy = \int_0^{x^2} 3\ dy = 3x^2$. Therefore, $f_X(0.5) = 3/4$.

Functions of multiple random variables

$Z = g(X, Y)$

Expected value rule:

$$\mathbf{E}[g(X,Y)] = \sum_x \sum_y g(x,y) p_{X,Y}(x,y) \qquad \mathbf{E}[g(X,Y)] = \int\int g(x,y) f_{X,Y}(x,y)\, dx\, dy$$

Linearity of expectations

$$\mathbf{E}[aX + b] = a\mathbf{E}[X] + b$$

$$\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$$

$$\boxed{\mathbf{E}[X_1 + \cdots + X_n] = \mathbf{E}[X_1] + \cdots + \mathbf{E}[X_n]}$$

Exercise: From joint PDFs to the marginals

2/5 points (graded)

For each one of the following formulas, identify those that are always true. All integrals are meant to be from $-\infty$ to ∞.

$$f_{X,Z}(a,b) = \int f_{X,Y,Z}(a', b, c)\ da'$$

No ✔ **Answer:** No

$$f_{X,Z}(a,c) = \int f_{X,Y,Z}(a, b, c)\ db$$

Yes ✔ **Answer:** Yes

$$f_{X,Z}(a,b) = \int f_{X,Y,Z}(a, b, c)\ dc$$

Yes ✘ **Answer:** No

$$f_Y(a) = \int\int\int f_{U,V,X,Y}(a, b, c, s)\ db\, dc\, ds$$

Yes ✘ **Answer:** No

$$f_Y(a) = \int\int\int f_{U,V,X,Y}(s, c, b, a)\ db\, dc\, ds$$

No ✘ **Answer:** Yes

Solution:

In each case, we need to "integrate out" the arguments associated with random variables that do not appear on the left-hand side. Thus, the correct formulas are:

$$f_{X,Z}(a,c) = \int f_{X,Y,Z}(a, b, c)\ db$$

and

$$f_Y(a) = \int\int\int f_{U,V,X,Y}(s, c, b, a)\ db\, dc\, ds.$$

The joint CDF

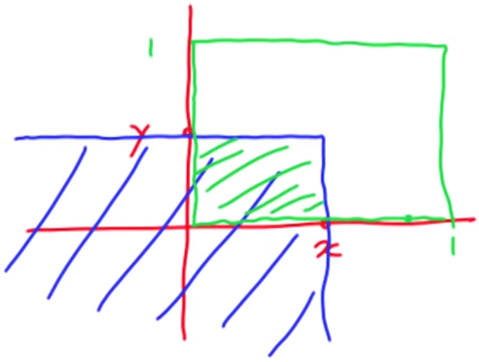
$$F_X(x) = \mathbf{P}(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt \qquad f_X(x) = \frac{dF_X}{dx}(x)$$

$$F_{X,Y}(x,y) = \mathbf{P}(X \le x, Y \le y) = \int_{-\infty}^{y}\left[\int_{-\infty}^{x} \underbrace{f_{X,Y}(s,t)}\,ds\right]dt$$

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x\,\partial y}(x,y)$$

$$F_{X,Y}(x,y) = xy$$

$$f_{X,Y}(x,y) = 1$$



So by differentiating the CDF with respect to y and then x (order doesn't matter) then we get the joint PDF

Exercise: Joint CDFs

3/3 points (graded)

a) Is it always true that if $x < x'$, then $F_{X,Y}(x,y) \le F_{X,Y}(x',y)$?

Yes ✔ **Answer:** Yes

b) Suppose that the random variables X and Y are jointly continuous and take values on the unit square, i.e., $0 \le x \le 1$ and $0 \le y \le 1$. Is $F_{X,Y}(x,y) = (x+2y)^2/9$ a legitimate joint CDF? *Hint:* Consider $F_{X,Y}(0,1)$.

No ✔ **Answer:** No

c) As above, suppose that the random variables X and Y are jointly continuous and take values on the unit square, i.e., $0 \le x \le 1$ and $0 \le y \le 1$. The joint CDF on that set is of the form $xy(x+y)/2$. Find an expression for the joint PDF which is valid for (x,y) in the unit square. Enter an algebraic function of x and y using standard notation.

x+y ✔ **Answer:** x+y

$x + y$

STANDARD NOTATION

Solution:

a) Since $x < x'$, the event $\{X \le x, Y \le y\}$ is a subset of the event $\{X \le x', Y \le y\}$, and therefore $F_{X,Y}(x,y) = \mathbf{P}(X \le x, Y \le y) \le \mathbf{P}(X \le x', Y \le y) = F_{X,Y}(x',y)$.

b) Since the random variables are nonnegative, we have $F_{X,Y}(0,1) = \mathbf{P}(X \le 0 \text{ and } Y \le 1) = \mathbf{P}(X = 0 \text{ and } Y \le 1) \le \mathbf{P}(X = 0) = 0$, where the last equality holds because X is a continuous random variable. But zero is different from $(0 + 2 \cdot 1)^2/9$. Therefore, we do not have a legitimate joint CDF.

c) The joint CDF is of the form $x^2y/2 + y^2x/2$. The partial derivative with respect to x is $xy + y^2/2$. Taking now the partial derivative with respect to y, we obtain $x + y$.

Conditioning on a random variable; Independence; Bayes' rule

Conditional PDFs, given another r.v.

$$p_{X|Y}(x \mid y) = \mathbf{P}(X = x \mid Y = y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}, \quad \text{if } p_Y(y) > 0$$

$p_{X,Y}(x,y)$	$f_{X,Y}(x,y)$
$p_{X|A}(x)$	$f_{X|A}(x)$
$p_{X|Y}(x \mid y)$	$f_{X|Y}(x \mid y)$

Definition: $$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \quad \text{if } f_Y(y) > 0$$

$$\mathbf{P}(x \le X \le x + \delta \mid \underbrace{A}_{Y=y \to Y\approx y}) \approx f_{X|A}(x) \cdot \delta, \qquad \text{where } \mathbf{P}(A) > 0$$

$$\mathbf{P}(x \le X \le x + \delta \mid y \le Y \le y + \epsilon) \approx \frac{f_{X,Y}(x,y)\,\delta\,\cancel{\epsilon}}{f_Y(y)\,\cancel{\epsilon}} = f_{X|Y}(x|y)\,\delta$$

Definition: $$\mathbf{P}(X \in A \mid Y = y) = \int_A f_{X\,|\,Y}(x \mid y)\,dx$$

can't look at event that Y=y as it is 0, therefore look at a small interval around y

Exercise: Conditional PDF

1/2 points (graded)

The random variables X and Y are jointly continuous, with a joint PDF of the form

$$f_{X,Y}(x,y) = \begin{cases} cxy, & \text{if } 0 \le x \le y \le 1, \\ 0, & \text{otherwise,} \end{cases}$$

where c is a normalizing constant.

a) Is it true that $f_{X|Y}(2 \mid 0.5)$ is equal to zero?

Yes ✔ **Answer:** Yes

b) Is it true that $f_{X|Y}(0.5 \mid 2)$ is equal to zero?

Yes ✖ **Answer:** No

Solution:

a) Values of Y around 0.5 have positive probability, so that $f_Y(0.5) > 0$, and $f_{X|Y}(2 \mid 0.5)$ is therefore well-defined. But $x = 2$ is outside the range of values of X, and $f_{X,Y}(2, 0.5) = 0$, from which it follows that $f_{X|Y}(2 \mid 0.5) = 0$.

b) Since $y = 2$ is outside the range of values of Y, we have $f_Y(2) = 0$, and the conditional PDF $f_{X|Y}(0.5 \mid 2)$ is undefined.

Comments on conditional PDFs

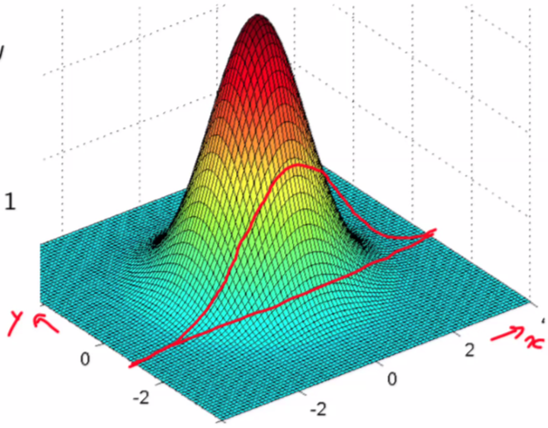
$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

- $f_{X|Y}(x \mid y) \geq 0$

- Think of value of Y as fixed at some y
  shape of $f_{X|Y}(\cdot \mid y)$: slice of the joint

- $\displaystyle\int_{-\infty}^{\infty} f_{X|Y}(x \mid y)\, dx = \frac{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx}{f_Y(y)} = 1$

- Multiplication rule:
  $f_{X,Y}(x,y) = f_Y(y) \cdot f_{X|Y}(x \mid y)$
  $\qquad = f_X(x) \cdot f_{Y|X}(y \mid x)$



Exercise: Conditional PDFs

1/2 points (graded)

The random variables X and Y are jointly continuous, with a joint PDF of the form

$$f_{X,Y}(x,y) = \begin{cases} cxy, & \text{if } 0 \leq x \leq y \leq 1, \\ 0, & \text{otherwise,} \end{cases}$$

where c is a normalizing constant.

For $x \in [0, 0.5]$, the conditional PDF $f_{X|Y}(x \mid 0.5)$ is of the form ax^b. Find a and b. Your answers should be numbers.

$a =$ 0.5 ✖ **Answer:** 8

$b =$ 1 ✔ **Answer:** 1

Solution:

We have $f_{X|Y}(x \mid 0.5) = \dfrac{f_{X,Y}(x, 0.5)}{f_Y(0.5)}$.

Having fixed $y = 0.5$, the conditional PDF is to be viewed as a function of x. For those values of x that are possible (i.e., $x \in [0, 0.5]$), the conditional PDF will be proportional to the joint PDF, hence of the form ax, for some constant a. This implies that $b = 1$. To find the normalizing constant, we use the normalization equation

$$1 = \int_0^{0.5} f_{X|Y}(x \mid 0.5)\ dx = \int_0^{0.5} ax\, dx = a \cdot \frac{x^2}{2}\Big|_0^{0.5} = \frac{a}{8},$$

which yields $a = 8$.

Total probability and expectation theorems

$$p_X(x) = \sum_y p_Y(y) p_{X|Y}(x\,|\,y) \qquad f_X(x) = \int_{-\infty}^{\infty} \overbrace{f_Y(y) f_{X|Y}(x\,|\,y)}^{f_{X,Y}(x,y)}\, dy \quad \text{Thm.}$$

$$\mathbf{E}[X \mid Y = y] = \sum_x x p_{X|Y}(x\,|\,y) \qquad \mathbf{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x\,|\,y)\, dx \quad \text{Def.}$$

$$\mathbf{E}[X] = \sum_y p_Y(y) \mathbf{E}[X \mid Y = y] \qquad \mathbf{E}[X] = \int_{-\infty}^{\infty} f_Y(y) \mathbf{E}[X \mid Y = y]\, dy$$

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

$$= \int_{-\infty}^{\infty} x \underbrace{\int_{-\infty}^{\infty} f_Y(y) f_{X|Y}(x|y)\, dy}\, dx$$

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

- Expected value rule...

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

Exercise: Expected value rule and total expectation theorem

8/8 points (graded)

Let X, Y, and Z be jointly continuous random variables. Assume that all conditional PDFs and expectations are well defined. E.g., when conditioning on $X = x$, assume that x is such that $f_X(x) > 0$. For each one of the following formulas, state whether it is true for all choices of the function g or false (i.e., not true for all choices of g).

1. $\mathbf{E}\big[g(Y) \mid X = x\big] = \int g(y) f_{Y|X}(y \mid x)\, dy$

   True ✔ **Answer:** True

2. $\mathbf{E}\big[g(y) \mid X = x\big] = \int g(y) f_{Y|X}(y \mid x)\, dy$

   False ✔ **Answer:** False

3. $\mathbf{E}\big[g(Y)\big] = \int \mathbf{E}\big[g(Y) \mid Z = z\big] f_Z(z)\, dz$

   True ✔ **Answer:** True

4. $\mathbf{E}\big[g(Y) \mid X = x, Z = z\big] = \int g(y) f_{Y|X,Z}(y \mid x, z)\, dy$

   True ✔ **Answer:** True

5. $\mathbf{E}\big[g(Y) \mid X = x\big] = \int \mathbf{E}\big[g(Y) \mid X = x, Z = z\big] f_{Z|X}(z \mid x)\, dz$

   True ✔ **Answer:** True

6. $\mathbf{E}\left[g(X,Y) \mid Y=y\right] = \mathbf{E}\left[g(X,y) \mid Y=y\right]$

True ✔ **Answer:** True

7. $\mathbf{E}\left[g(X,Y) \mid Y=y\right] = \mathbf{E}\left[g(X,y)\right]$

False ✔ **Answer:** False

8. $\mathbf{E}\left[g(X,Z) \mid Y=y\right] = \int g(x,z)\, f_{X,Z|Y}(x,z \mid y)\, dy$

False ✔ **Answer:** False

Solution:

1. True. This is the usual expected value rule, applied to a conditional model where we are given that $X = x$.
2. False. Here the quantity inside the expectation, $g(y)$, is a number (not a random variable). The left-hand side is a function of y, whereas on the right-hand side, y, is a dummy variable that gets integrated away. So, the formula is wrong on a purely syntactical basis (the left-hand side depends on y, while the right-hand side does not).
3. True. This is the total expectation theorem, where we condition on the events $Z = z$.
4. True. This is the usual expected value rule, applied to a conditional model where we are given that $X = x$ and $Z = z$.
5. True. This is the same total expectation theorem as in the third part, except that everything is calculated within a conditional model in which event $X = x$ is known to have occurred.
6. True. When we condition on $Y = y$, we know the value of Y, and we can replace $g(X,Y)$ by $g(X,y)$.
7. False. Given that $Y = y$, we need to somehow take into account the conditional distribution of X, whereas the right-hand side is determined by the unconditional PDF of X.
8. False. The left-hand side is a function of y, whereas the right-hand side (after y is integrated out) is a function of x and z. The correct form (expected value rule, in a conditional model) is:

$$\mathbf{E}\left[g(X,Z) \mid Y=y\right] = \int\int g(x,z)\, f_{X,Z|Y}(x,z \mid y)\, dx\, dz.$$

Independence

$p_{X,Y}(x,y) = p_X(x)\,p_Y(y)$, for all x, y

$$\boxed{f_{X,Y}(x,y) = f_X(x)\,f_Y(y), \quad \text{for all } x \text{ and } y}$$

$f_{Y|X} = f_Y$

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

- equivalent to: $f_{X|Y}(x\,|\,y) = f_X(x)$, for all y with $f_Y(y) > 0$ and all x

If X, Y are **independent:** $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$

$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$

$g(X)$ and $h(Y)$ are also independent: $\mathbf{E}[g(X)h(Y)] = \mathbf{E}[g(X)] \cdot \mathbf{E}[h(Y)]$

Exercise: Definition of independence

1/1 point (graded)

Suppose that X and Y are independent, with a joint PDF that is uniform on a certain set S: $f_{X,Y}(x,y)$ is constant on S, and zero otherwise. The set S

- ○ must be a square.
- ● must be a set of the form $\{(x,y) : x \in A,\ y \in B\}$ (known as the Cartesian product of two sets A and B).
- ○ can be any set.

✔

Solution:

Let A be the set of all x on which $f_X(x)$ is positive and let B be the set of all y on which $f_Y(y)$ is positive. Then, the set S, on which $f_{X,Y}(x,y) = f_X(x)\,f_Y(y) > 0$, will be the Cartesian product of A with B; it is not necessarily a square, but it cannot be an arbitrary set.

Exercise: Independence and expectations II

2/3 points (graded)

Let X, Y, and Z be independent jointly continuous random variables, and let g, h, r be some functions. For each one of the following formulas, state whether it is true for all choices of the functions g, h, and r, or false (i.e., not true for all choices of these functions). Do not attempt formal derivations; use an intuitive argument.

1. $\mathbf{E}\big[g(X,Y)\,h(Z)\big] = \mathbf{E}\big[g(X,Y)\big]\cdot\mathbf{E}\big[h(Z)\big]$

   True ✔ **Answer:** True

2. $\mathbf{E}\big[g(X,Y)\,h(Y,Z)\big] = \mathbf{E}\big[g(X,Y)\big]\cdot\mathbf{E}\big[h(Y,Z)\big]$

   True ✖ **Answer:** False

3. $\mathbf{E}\big[g(X)\,r(Y)\,h(Z)\big] = \mathbf{E}\big[g(X)\big]\cdot\mathbf{E}\big[r(Y)\big]\cdot\mathbf{E}\big[h(Z)\big]$

   True ✔ **Answer:** True

Solution:

1. True. Using our intuitive understanding of independence, the pair of random variables (X,Y) does not provide any information on Z. Therefore, (X,Y) and Z are independent. It follows that $g(X,Y)$ and $h(Z)$ are independent, from which the formula follows.

2. False. The random variable Y appears in both functions g and h, so that $g(X,Y)$ and $h(Y,Z)$ will be, in general, dependent. For an example, suppose that $g(X,Y) = h(Y,Z) = Y$, in which case the statement becomes $\mathbf{E}[Y^2] = \big(\mathbf{E}[Y]\big)^2$, which we know to be false in general.

3. True. Using the first part, and then again the independence of X with Y, we have $\mathbf{E}\big[g(X)\,r(Y)\,h(Z)\big] = \mathbf{E}\big[g(X)\,r(Y)\big]\cdot\mathbf{E}\big[h(Z)\big] = \mathbf{E}\big[g(X)\big]\cdot\mathbf{E}\big[r(Y)\big]\cdot\mathbf{E}\big[h(Z)\big]$.

a) Suppose that X and Y are independent. Is it true that their joint CDF satisfies $F_{X,Y}(x,y) = F_X(x)\,F_Y(y)$, for all x and y?

Yes ✔ **Answer:** Yes

b) Suppose that $F_{X,Y}(x,y) = F_X(x)\,F_Y(y)$, for all x and y. Is it true that X and Y are independent?
Hint: Recall the formula $f_{X,Y}(x,y) = (\partial^2/\partial x \partial y)\,F_{X,Y}(x,y)$.

No ✖ **Answer:** Yes

Solution:

a) Yes. We have

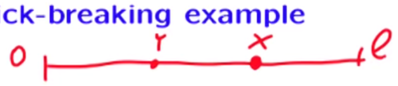
$$\begin{aligned}
F_{X,Y}(x,y) &= \mathbf{P}(X \le x, Y \le y) \\
&= \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(x,y)\; dx\, dy \\
&= \int_{-\infty}^{x} f_X(x)\; dx \int_{-\infty}^{y} f_Y(y)\; dy \\
&= F_X(x)\,F_Y(y)\,.
\end{aligned}$$

b) True. Using the formula in the hint, we find that

$$\begin{aligned}
f_{X,Y}(x,y) &= \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y) \\
&= \frac{\partial^2}{\partial x \partial y} F_X(x)\,F_Y(y) \\
&= \frac{\partial}{\partial x} F_X(x)\,\frac{\partial}{\partial y} F_Y(y) \\
&= f_X(x)\,f_Y(y)\,,
\end{aligned}$$

and therefore we have independence.

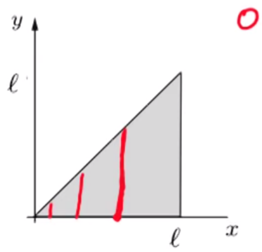
Stick-breaking example

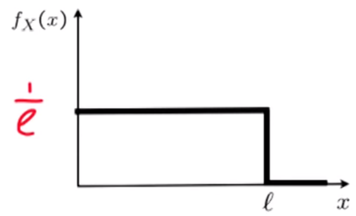


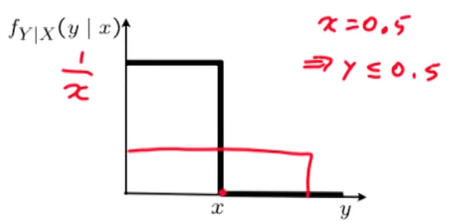
- Break a stick of length ℓ twice
  - first break at X: uniform in $[0,\ell]$
  - second break at Y: uniform in $[0,X]$

$$f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y \mid x) = \frac{1}{\ell x}$$

$$0 \le y \le x \le \ell$$

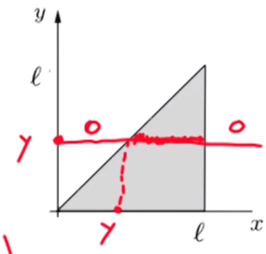






Y is dependent on X because the stick is broken at X first and determines Ymax

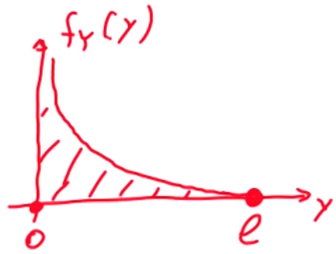
Stick-breaking example



$$f_{X,Y}(x,y) = \frac{1}{\ell x}, \qquad 0 \le y \le x \le \ell$$

$$f_Y(y) = \int f_{X,Y}(x,y)\,dx = \int_y^{\ell} \frac{1}{\ell x}\,dx = \frac{1}{\ell}\log\left(\frac{\ell}{y}\right)$$

$$\mathbf{E}[Y] = \int_0^{\ell} y\,\frac{1}{\ell}\log\left(\frac{\ell}{y}\right)dy$$



- Using total expectation theorem:

$$E[Y] = \int_0^{\ell} \frac{1}{\ell}\, E[Y|X=x]\,dx = \int_0^{\ell} \underbrace{\frac{1}{\ell}}_{f_X(x)} \frac{x}{2}\,dx = \frac{1}{2}E[X] = \frac{1}{2}\cdot\frac{\ell}{2} = \frac{\ell}{4}.$$

from y to L because the slope is 1
E[Y] = L/4 makes sense as we expect the expected value to half if we snap the stick once so L/2 and if we snap it again then it becomes L/4

Exercise: Stick-breaking

3/3 points (graded)

Consider the same stick-breaking problem as in the previous clip, and let $\ell = 1$. Recall that $f_{X,Y}(x,y) = 1/x$ when $0 \le y \le x \le 1$.

a) Conditioned on $Y = 2/3$, the conditional PDF of X is nonzero when $a \le x \le b$. Find a and b.

$a =$ 2/3 ✔ **Answer:** 0.66667

$b =$ 1 ✔ **Answer:** 1

b) On the range found in part (a), the conditional PDF $f_{X|Y}(x \mid 2/3)$ is of the form cx^d for some constants c and d. Find d.

$d =$ -1 ✔ **Answer:** -1

Solution:

a) Since the joint PDF is nonzero only for $0 \le y \le x \le 1$, it follows that given that $Y = 2/3$, X ranges on the interval $[2/3, 1]$.

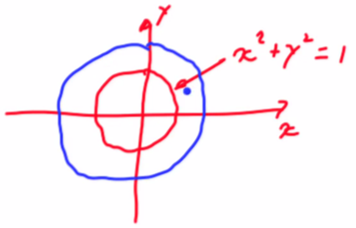
b) As a function of x, the conditional PDF has the same functional form (within a normalizing constant) as the joint PDF, and so it is of the form c/x, from which we conclude that $d = -1$.

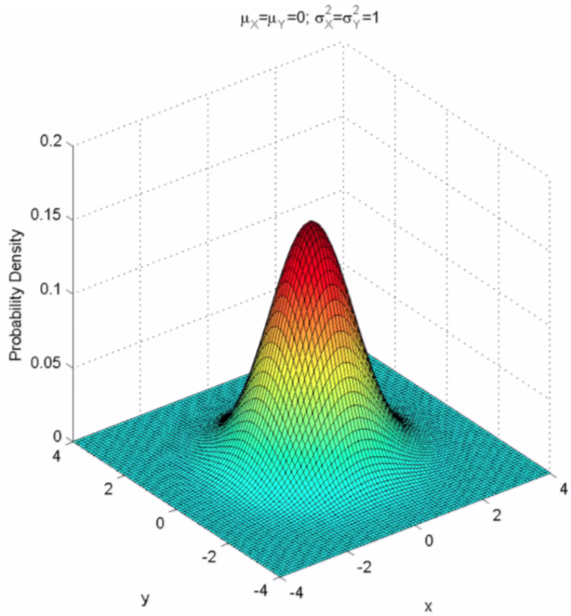
Independent standard normals

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

$$= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\}$$

$$= \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(x^2+y^2)\right\}$$



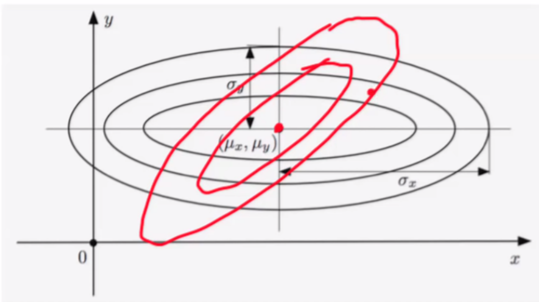


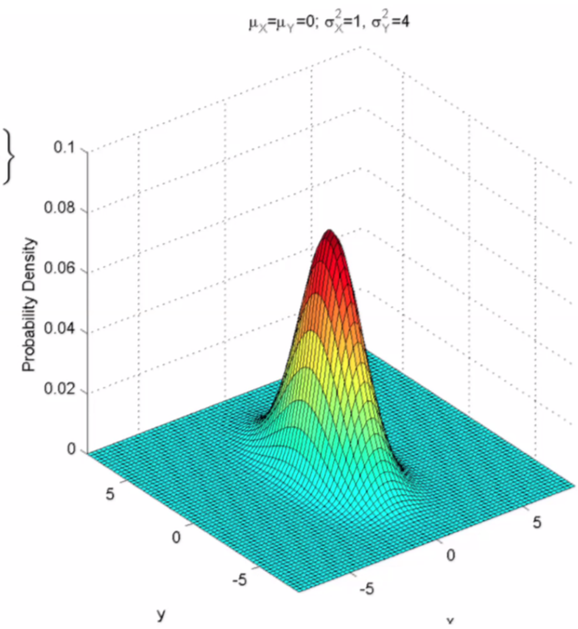
On each circle the joint PDF is a constant

Independent normals

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

$$= \underbrace{\frac{1}{2\pi\sigma_x\sigma_y}} \exp\left\{-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2}\right\}$$





First part is a normalising constant so the joint PDF integrates to 1 (1/2.pi.sigx.sigy)
Joint PDF is largest when x=mu(x) and y =mu(y)
Y has a larger variance than X so it appears stretched out
The centre of the bell is determined by the means
If we had ‘diagonal’ ellipsis as has been sketched then we would have dependence between the two variables

Exercise: Independent normals

2/4 points (graded)

The random variables X and Y have a joint PDF of the form $f_{X,Y}(x,y) = c \cdot \exp\left\{-\frac{1}{2}\left(4x^2 - 8x + y^2 - 6y + 13\right)\right\}$.

$\mathbf{E}[X] =$ 1 ✔ **Answer:** 1

$\mathsf{Var}(X) =$ 4 ✖ **Answer:** 0.25

$\mathbf{E}[Y] =$ 3 ✔ **Answer:** 3

$\mathsf{Var}(Y) =$ 4 ✖ **Answer:** 1

Solution:

We rewrite the joint PDF in the form

$$f_{X,Y}(x,y) = c \cdot \exp\left\{-\frac{1}{2}\left(\frac{(x-1)^2}{1/4} + (y-3)^2\right)\right\},$$

and we recognize that we are dealing with the joint PDF of two independent normals with $\mathbf{E}[X] = 1$, $\mathsf{Var}(X) = 1/4$, $\mathbf{E}[Y] = 3$, and $\mathsf{Var}(Y) = 1$.

Bayes rule variations

The Bayes rule — a theme with variations

Inference

$p_{X|Y}(\cdot \mid y)$

$$X \longrightarrow Y$$

unobserved value x | observed value y

prior $p_X(\cdot)$ | model $p_{Y|X}(\cdot \mid \cdot)$

$$\begin{aligned} p_{X,Y}(x,y) &= p_X(x)\, p_{Y|X}(y \mid x) \\ &= p_Y(y)\, p_{X|Y}(x \mid y) \end{aligned} \qquad \begin{aligned} f_{X,Y}(x,y) &= f_X(x)\, f_{Y|X}(y \mid x) \\ &= f_Y(y)\, f_{X|Y}(x \mid y) \end{aligned}$$

$$p_{X|Y}(x \mid y) = \frac{p_X(x)\, p_{Y|X}(y \mid x)}{p_Y(y)} \qquad f_{X|Y}(x \mid y) = \frac{f_X(x)\, f_{Y|X}(y \mid x)}{f_Y(y)}$$

posterior

$$p_Y(y) = \sum_{x'} p_X(x')\, p_{Y|X}(y \mid x') \qquad f_Y(y) = \int f_X(x')\, f_{Y|X}(y \mid x')\, dx'$$

x′ is a dummy variable, could also just be z for example
it is not the derivative of x

Exercise: The discrete Bayes rule

0/1 point (graded)

The bias of a coin (i.e., the probability of Heads) can take three possible values, $1/4$, $1/2$, or $3/4$, and is modeled as a discrete random variable Q with PMF

$$p_Q(q) = \begin{cases} 1/6, & \text{if } q = 1/4, \\ 2/6, & \text{if } q = 2/4, \\ 3/6, & \text{if } q = 3/4, \\ 0, & \text{otherwise.} \end{cases}$$

Let K be the total number of Heads in two independent tosses of the coin. Find $p_{Q|K}(3/4 \mid 2)$.

1/3

✖ **Answer:** 0.75

Solution:

The Bayes rule for discrete random variables gives

$$p_{Q|K}(3/4 \mid 2) = \frac{p_Q(3/4)\, p_{K|Q}(2 \mid 3/4)}{p_K(2)} = \frac{(3/6)\cdot(3/4)^2}{p_K(2)} = \frac{(3/6)\cdot(3/4)^2}{3/8} = \frac{3}{4}.$$

To find $p_K(2)$, we used the total probability theorem:

$$p_K(2) = \sum_q p_Q(q)\, p_{K|Q}(2 \mid q) = (1/6)\cdot(1/4)^2 + (2/6)\cdot(2/4)^2 + (3/6)\cdot(3/4)^2 = 3/8.$$

The Bayes rule — one discrete and one continuous random variable

K: discrete $\qquad$ Y: continuous

$$P(K=k,\ y \le Y \le y+\delta) \qquad \delta > 0,\ \delta \approx 0$$

$$= P(K=k)\, P(y \le Y \le y+\delta \mid K=k) \approx p_K(k)\, f_{Y|K}(y \mid k)\, \cancel{\delta}$$

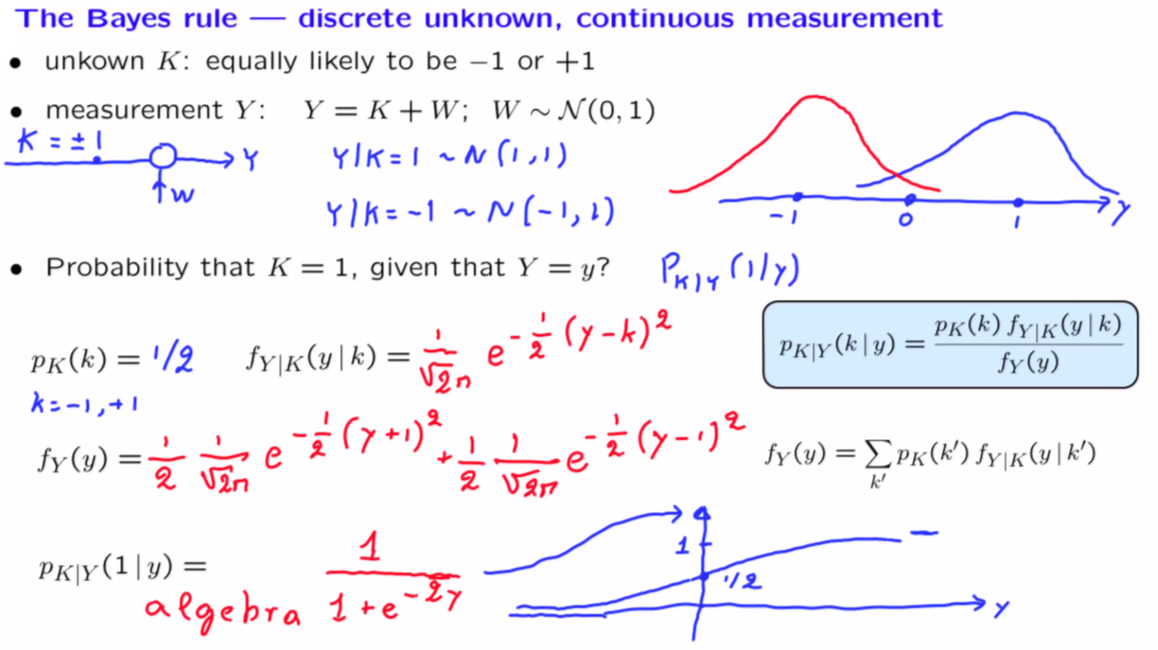
$$= P(y \le Y \le y+\delta)\, P(K=k \mid y \le Y \le y+\delta) \approx f_Y(y)\, \cancel{\delta}\, p_{K|Y}(k \mid y)$$

$$\boxed{p_{K|Y}(k \mid y) = \frac{p_K(k)\, f_{Y|K}(y \mid k)}{f_Y(y)}} \qquad \boxed{f_{Y|K}(y \mid k) = \frac{f_Y(y)\, p_{K|Y}(k \mid y)}{p_K(k)}}$$

$$f_Y(y) = \sum_{k'} p_K(k')\, f_{Y|K}(y \mid k') \qquad p_K(k) = \int f_Y(y')\, p_{K|Y}(k \mid y')\, dy'$$

This version is useful if we have a continuous noisy signal that is obscuring a discrete variable i.e. inferring a discrete random variable from a continuous random variable and vice versa

Example, detection of a binary signal



This is a prototype simple model of analysis for signal and noise in communications
After getting all parts of the formula a lot of algebra gives a very simple result and plotting it shows that if the sum of noise and signal (Y) is very positive then the signal was likely +1 (and vice versa for negative).
But if Y is 0 then there’s an equal chance of the signal being +1 and -1
f(Y|K) is given by the fact that the distribution of Y is normal with deviation 1

Exercise: Discrete unknown, continuous measurement

1/1 point (graded)

Let K be a discrete random variable that can take the values $1, 2$, and 3, all with equal probability. Suppose that X takes values in $[0, 1]$ and that for x in that interval we have

$$f_{X|K}(x \mid k) = \begin{cases} 1, & \text{if } k = 1, \\ 2x, & \text{if } k = 2, \\ 3x^2, & \text{if } k = 3. \end{cases}$$

Find the probability that $K = 1$, given that $X = 1/2$.

12/33 ✔ **Answer:** 0.36364

Solution:

Using the appropriate form of the Bayes rule, we have

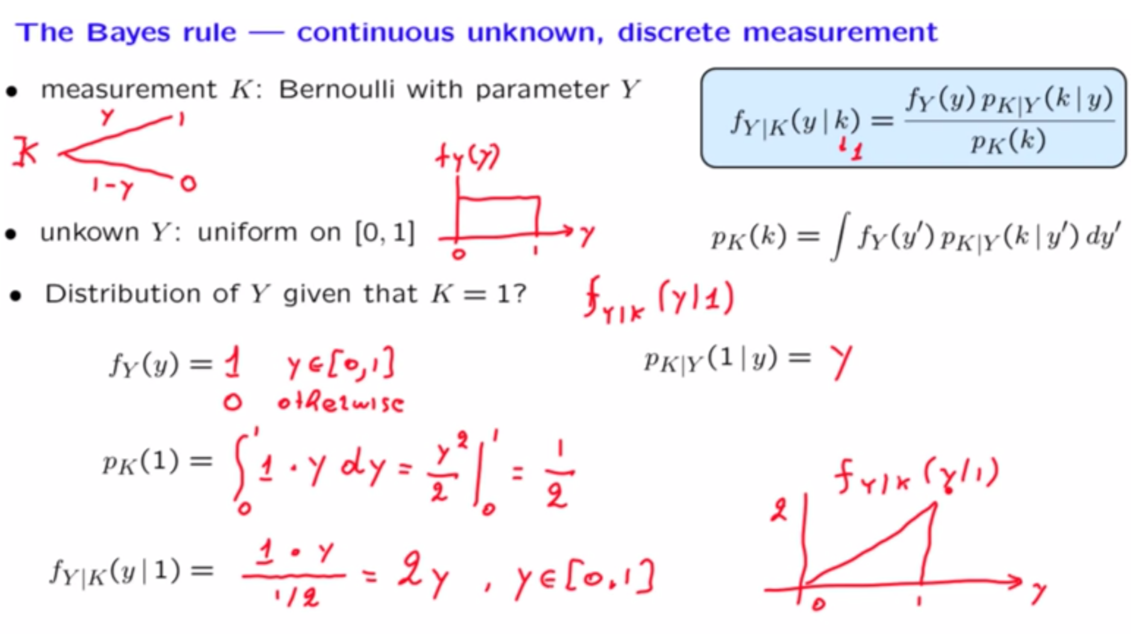
$$p_{K|X}(1 \mid 1/2) = \frac{p_K(1)\, f_{X|K}(1/2 \mid 1)}{f_X(1/2)} = \frac{(1/3) \cdot 1}{f_X(1/2)} = \frac{1/3}{11/12} = 4/11.$$

To find $f_X(1/2)$, we used the total probability theorem:

$$\begin{aligned} f_X(1/2) &= \sum_k p_K(k)\, f_{X|K}(1/2 \mid k) \\ &= (1/3) \cdot 1 + (1/3) \cdot (2 \cdot (1/2)) + (1/3) \cdot (3 \cdot (1/2)^2) \\ &= 11/12. \end{aligned}$$

Inference of the bias of a coin

Trying to guess the bias of a coin given the amount of heads that are turning up



This tells us that once we observe 1/1 heads then it’s possible that the coin is more biased to heads

Exercise: Inference of the bias of a coin

1/1 point (graded)

The random variable K is geometric with a parameter which is itself a uniform random variable Q on $[0, 1]$. Find the value $f_{Q|K}(0.5 \mid 1)$ of the conditional PDF of Q, given that $K = 1$. *Hint:* Use the result in the last segment.

1 **Answer:** 1

Solution:

We identify Q with the variable Y in the last segment. The information that $K = 1$ is the information that the first coin flip resulted in Heads, which is the same as the information that $K = 1$ in the last segment. Therefore, the conditional PDF of Q is $2q$, which for $q = 0.5$ evaluates to 1.