Lecture 6b

ECE 361
Probability for Engineers
Fall, 2016
Steven Weber

Tuesday November 1, 2016

DREXEL UNIVERSITY
Electrical and Computer Engineering
College of Engineering

Outline

The Gaussian distribution

Gaussian PDF. The Gaussian (normal) RV X has support $\mathcal{X} = \mathbb{R}$ and PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}, \; x \in \mathbb{R}.$$

This distribution has **two** free parameters: $\mu \in \mathbb{R}$ and $\sigma \geq 0$

- $\mu \in \mathbb{R}$. We will show $\mathbb{E}[X] = \mu$. Changing μ **shifts** the PDF so that it is centered at μ. The **standard** value of μ is 0.
- $\sigma \in \mathbb{R}_+$. We will show $\mathrm{Var}[X] = \sigma^2$. Changing σ **scales** the PDF, either stretching it (for $\sigma > 1$) or compressing it (for $\sigma \in (0,1)$). The **standard** value of σ is 1.

We write $X \sim N(\mu, \sigma)$ to denote that the RV is normally distributed with parameters (μ, σ). Other authors write $X \sim N(\mu, \sigma^2)$.

The standard Gaussian distribution

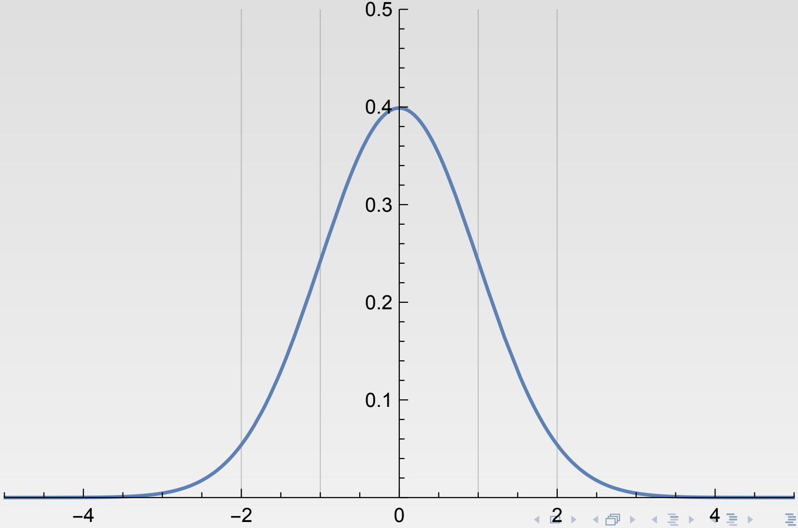
Gaussian PDF. The **standard** Gaussian (normal) RV Z has support $\mathcal{Z} = \mathbb{R}$ and PDF

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}}, \; z \in \mathbb{R}.$$

This corresponds to a normal RV with $\mu = 0$ and $\sigma = 1$. A standard normal is often denoted $Z \sim N(0,1)$.

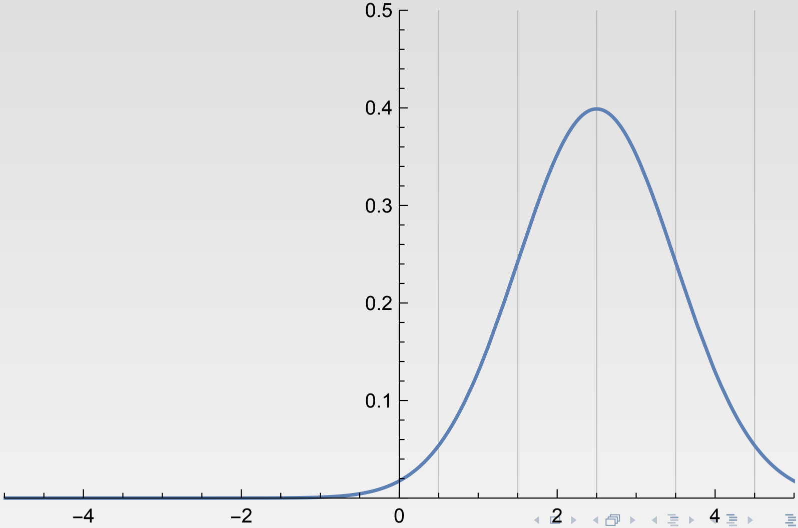
The Gaussian distribution

The $N(\mu,\sigma)$ distribution: $\mu = 0$ and $\sigma = 1$. Gridlines at $\mu, \pm\sigma, \pm 2\sigma$.



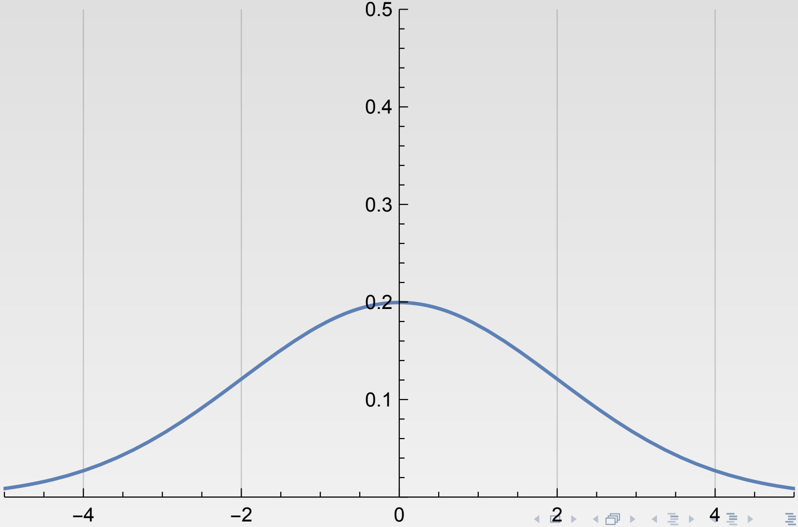
The Gaussian distribution

The $N(\mu, \sigma)$ distribution: $\mu = 2.5$ and $\sigma = 1$. Gridlines at $\mu, \pm\sigma, \pm 2\sigma$.



The Gaussian distribution

The $N(\mu, \sigma)$ distribution: $\mu = 0$ and $\sigma = 2$. Gridlines at $\mu, \pm\sigma, \pm 2\sigma$.



The Gaussian distribution

To be a valid PDF, we must show that $f_X(x)$ is nonnegative and integrates to one:

- Nonnegativity is obvious.
- Integration to one is **not** obvious, and is in fact **difficult** to show:

$$\int_{-\infty}^{\infty} f_X(x) \mathrm{d}x = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x = 1.$$

But is nonetheless **true**.

The Gaussian distribution

We know that given an **arbitrary** PDF $f_X(x)$ we can obtain its CDF $F_X(x)$ by integration:

$$F_X(x) = \int_{-\infty}^{x} f_X(t) \mathrm{d}t.$$

We **hope** that we can **solve** this integral to provide an **explicit** expression for $F_X(x)$.

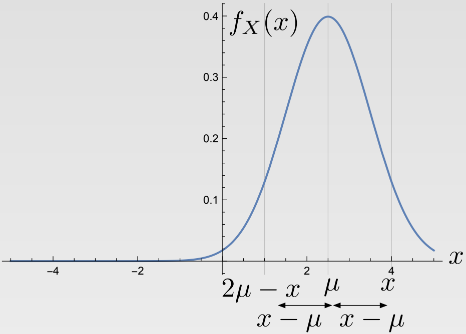
For the normal distribution we have:

$$F_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} \mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}} \mathrm{d}t.$$

It is not easy, but it can be shown that it is **not possible** to solve this integral explicitly. The CDF for the normal distribution is **only** expressible as an integral of the PDF. The normal CDF must be found by a computer or through tables.

The Gaussian distribution

The normal distribution is **symmetric** around the parameter μ: $f_X(x) = f_X(2\mu - x)$.



To see this, observe:

$$f_X(2\mu - x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{((2\mu - x) - \mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\mu - x)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x - \mu)^2}{2\sigma^2}} = f_X(x).$$

The Gaussian distribution

Any distribution that is symmetric about some number μ has two properties:

- μ is the **median**: $\mathbb{P}(X > \mu) = \mathbb{P}(X \leq \mu) = 1/2$
- μ is the **mean**: $\mathbb{E}[X] = \mu$.

Proof of the median: split the integral at μ and use a change of variable:

$$\begin{aligned} 1 &= \int_{-\infty}^{\mu} f_X(x)\mathrm{d}x + \int_{\mu}^{\infty} f_X(x)\mathrm{d}x \\ &= \int_{-\infty}^{\mu} f_X(2\mu - x)\mathrm{d}x + \mathbb{P}(X > \mu) \\ &= \int_{\mu}^{\infty} f_X(y)\mathrm{d}y + \mathbb{P}(X > \mu) \\ &= 2\mathbb{P}(X > \mu). \end{aligned}$$

The Gaussian distribution

Any distribution that is symmetric about some number μ has two properties:

- μ is the **median**: $\mathbb{P}(X > \mu) = \mathbb{P}(X \leq \mu) = 1/2$
- μ is the **mean**: $\mathbb{E}[X] = \mu$.

Proof of the mean: split the integral at μ and use a change of variable:

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\mu} x f_X(x)\mathrm{d}x + \int_{\mu}^{\infty} x f_X(x)\mathrm{d}x \\
&= \int_{\mu}^{\infty} (2\mu - y) f_X(y)\mathrm{d}y + \int_{\mu}^{\infty} x f_X(x)\mathrm{d}x \\
&= 2\mu \int_{\mu}^{\infty} f_X(y)\mathrm{d}y = 2\mu\mathbb{P}(X > \mu) = \mu.
\end{aligned}$$

The Gaussian distribution

The variance of **any** distribution is computed from the equation:

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x) \mathrm{d}x.$$

Use this equation to compute the variance of the normal distribution via the change of variable $y = (x-\mu)/\sigma$ and integration by parts:

$$\begin{aligned}
\mathrm{Var}(X) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 \mathrm{e}^{-\frac{y^2}{2}} \mathrm{d}y \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left(-y\mathrm{e}^{-\frac{y^2}{2}} \right) \Big|_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{y^2}{2}} \mathrm{d}y \\
&= \sigma^2.
\end{aligned}$$

The Gaussian distribution

Families of distributions we've seen thus far include:

- Uniform: $X \sim \mathrm{uni}([l,u])$, for $l < u$
- Exponential: $X \sim \exp(\lambda)$, for $\lambda > 0$
- Normal: $X \sim N(\mu, \sigma)$, for $\mu \in \mathbb{R}$ and $\sigma > 0$

Given two RVs from a given family, e.g., X, Y each uniform over $[c, d]$, it is **not** true in general that a linear combination of X, Y, e.g., $aX + bY$, is uniform.

- If X, Y are independent uniform RVs then $X + Y$ has the Irwin-Hall distribution
- If X, Y are independent exponential RVs then $X + Y$ has the gamma distribution

Most distribution families used in probability are not **closed** (also called **stable**) under linear combinations.

The Gaussian distribution

- The normal distribution, however, **is closed** (stable): linear combinations of normally distributed RVs are themselves normal.
- If $X \sim N(\mu_x, \sigma_x)$ and $Y \sim N(\mu_y, \sigma_y)$, then $aX + bY \sim N(\mu_z, \sigma_z)$ for some (μ_z, σ_z). We will not prove this fact.
- What are (μ_z, σ_z)?
    - Recall linearity of expectation: $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.
    - Recall variance for linear combinations of **independent** RVs: $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y)$.
    - Thus $\mu_z = a\mu_x + b\mu_y$ and $\sigma_z^2 = a^2\sigma_x^2 + b^2\sigma_y^2$.

The Gaussian distribution

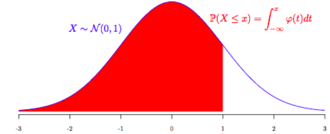
Recall that $Z \sim N(0,1)$ (with $\mu = 0$ and $\sigma = 1$) is the **standard** normal.

Standard normal PDF. The standard Gaussian (normal) RV Z has support $\mathcal{Z} = \mathbb{R}$ and PDF and CDF:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}}, \; F_Z(z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x, \; z \in \mathbb{R}.$$

- The value of $F_Z(z)$ is tabulated, with a sample table on the next slide.
- The value of $F_Z(z)$ is also available by computer for many computer packages:
  - Matlab: normcdf(z)
  - Mathematica: CDF[NormalDistribution[0, 1], z]
  - Python: from scipy.stats import norm; norm.cdf(z)

The Gaussian distribution



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

http://f.hypotheses.org/wp-content/blogs.dir/253/files/2013/10/Capture-d?cran-2013-10-15--14.22.40.png

The Gaussian distribution

The table on the previous slide gives the CDF $\Phi(z)$ for a **standard** normal, $Z \sim \Phi(0,1)$. But what if you wish to evaluate $F_X(x)$ for $X \sim N(\mu, \sigma)$? Answer: standardize.

- For **any** RV X, not just a normal RV. If X has expectation $\mathbb{E}[X]$ and standard deviation $\mathrm{Std}(X)$, then its standardized version is:

$$Y = \frac{X - \mathbb{E}[X]}{\mathrm{Std}(X)}.$$

- The RV Y has mean 0 and variance 1:

$$\begin{aligned} \mathbb{E}[Y] &= \mathbb{E}\left[\frac{X - \mathbb{E}[X]}{\mathrm{Std}(X)}\right] = \frac{\mathbb{E}[X - \mathbb{E}[X]]}{\mathrm{Std}(X)} = 0 \\ \mathrm{Var}(Y) &= \mathrm{Var}\left(\frac{X - \mathbb{E}[X]}{\mathrm{Std}(X)}\right) = \frac{\mathrm{Var}(\tilde{X})}{\mathrm{Var}(\tilde{X})} = 1. \end{aligned}$$

The Gaussian distribution

Therefore, if $X \sim N(\mu, \sigma)$ and we wish to evaluate $F_X(x)$ for some $x \in \mathbb{R}$:

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) \\ &= \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) \\ &= \mathbb{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) \\ &= F_Z\left(\frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right). \end{aligned}$$

We have standardized X into Z, and expressed $F_X(x)$ in terms of $\Phi(\cdot)$. The key insight is, for any $a > 0$:

$$\{X \leq x\} \Leftrightarrow \{X - a \leq x - a\}, \ \{X \leq x\} \Leftrightarrow \{X/a \leq x/a\}.$$

The Gaussian distribution

Example. The annual snowfall is $X \sim N(\mu, \sigma)$ with $\mu = 60$ inches and $\sigma = 20$ inches. Find the probability the snowfall will be at least $x = 80$ inches. Note $(x - \mu)/\sigma = 1$.

$$
\begin{aligned}
\mathbb{P}(X > x) &= \mathbb{P}\left(\frac{X - \mu}{\sigma} > \frac{x - \mu}{\sigma}\right) \\
&= \mathbb{P}(Y > 1) \\
&= 1 - \mathbb{P}(Y \leq 1) \\
&= 1 - \Phi(1) \\
&\approx 1 - 0.8413 = 0.1587.
\end{aligned}
$$

Outline

Outline

Joint PDFs of multiple RVs

The PDF f_X for a RV X obeys

$$\mathbb{P}(X \in B) = \int_B f_X(x)\mathrm{d}x,$$

for any $B \subseteq \mathbb{R}$. The PDF is normalized:

$$\int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x = 1.$$

Recall that the PDF f_X for a RV X is "probability per unit length":

$$\mathbb{P}(X \in [x, x+\delta]) = \int_x^{x+\delta} f_X(t)\mathrm{d}t \approx f_X(x)\delta \Rightarrow f_X(x) \approx \frac{\mathbb{P}(X \in [x, x+\delta])}{\delta}.$$

Joint PDFs of multiple RVs

Define a pair of RVs (X, Y) as jointly continuous if there exists a function $f_{X,Y}(x, y)$ (the joint PDF) such that

$$\mathbb{P}((X, Y) \in B) = \int_{(x,y)\in B} f_{X,Y}(x, y)\mathrm{d}x\mathrm{d}y$$

for all $B \subset \mathbb{R}^2$. Letting $B = \mathbb{R}^2$ we see the joint PDF must obey the normalization property:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\mathrm{d}x\mathrm{d}y = 1.$$

Consider some point (a, c) and some small δ and form the set $B = [a, a + \delta] \times [c, c + \delta]$ with area δ^2. Then

$$\mathbb{P}((X, Y) \in B) = \int_{c}^{c+\delta}\int_{a}^{a+\delta} f_{X,Y}(x, y)\mathrm{d}x\mathrm{d}y = f_{X,Y}(a, c)\delta^2,$$

so $f_{X,Y}(x, y)$ is the “probability per unit area” at point (x, y).

Joint PDFs of multiple RVs

Recall. Let (U,V) be discrete RVs with joint support $\mathcal{A}$ and joint PMF $p_{U,V}(u,v) = \mathbb{P}(U=u, V=v)$ obeying

$$\sum_{(u,v)\in\mathcal{A}} p_{U,V}(u,v) = 1.$$

Recall that we can obtain the marginal PMFs by summing over the "other" variable:

$$\begin{aligned} p_U(u) &= \sum_{v\in\mathcal{V}} p_{U,V}(u,v), \;\; u \in \mathcal{U} \\ p_V(v) &= \sum_{u\in\mathcal{U}} p_{U,V}(u,v), \;\; v \in \mathcal{V} \end{aligned}$$

Joint PDFs of multiple RVs

Back to the continuous case. Let (X, Y) have joint PDF $f_{X,Y}(x, y)$. For any $A \subset \mathbb{R}$ we have

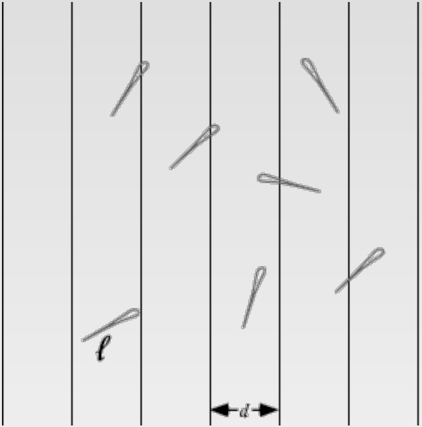
$$\mathbb{P}(X \in A) = \mathbb{P}(X \in A \cap Y \in \mathbb{R}) = \int_A \int_{-\infty}^{\infty} f_{X,Y}(x, y)\mathrm{d}y\mathrm{d}x.$$

Thus the marginal PDFs are:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\mathrm{d}y, \ \ f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\mathrm{d}x.$$

Joint PDFs of multiple RVs

Example. Buffon's needle. Lines separation d, needles length $l < d$.

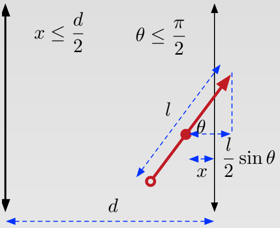


What is the probability a needle dropped on this surface intersects a line?

http://mathworld.wolfram.com/BuffonsNeedleProblem.html

Joint PDFs of multiple RVs

Example. Buffon's needle.



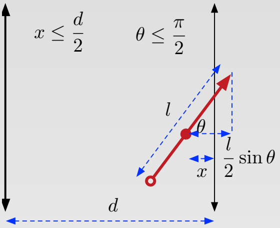
- x: distance from the center of the needle to the nearest line: $x \leq d/2$
- θ: acute angle made by the angle relative to the horizontal: $\theta \leq \pi/2$
- For a random needle: $X \sim \mathrm{uni}([0, d/2])$ and $\Theta \sim \mathrm{uni}([0, \pi/2])$
- (X, θ) are independent with uniform joint distribution:

$$f_{X,\Theta}(x, \theta) = f_X(x) f_\Theta(\theta) = \frac{2}{d} \times \frac{2}{\pi} = \frac{4}{\pi d},$$

for any $(x, d) \in [0, d/2] \times [0, \pi/2]$.

Joint PDFs of multiple RVs

Example. Buffon's needle.



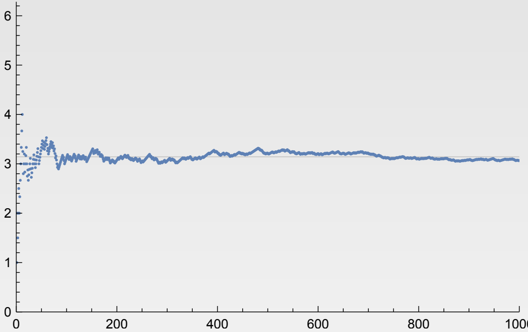
- The event the needle hits the line is the event $X \leq \frac{l}{2} \sin \Theta$ and thus:

$$\begin{aligned} \mathbb{P}(X \leq \frac{l}{2} \sin \Theta) &= \int_{(x,\theta): x \leq \frac{l}{2} \sin\theta} f_{X,\Theta}(x,\theta) \mathrm{d}x\mathrm{d}\theta \\ &= \frac{4}{\pi d} \int_0^{\pi/2} \int_0^{(l/2)\sin\theta} \mathrm{d}x\mathrm{d}\theta = \frac{2l}{\pi d}. \end{aligned}$$

Joint PDFs of multiple RVs

Example. Buffon's needle.

- We showed $p = \mathbb{P}(\text{hit}) = \frac{2l}{\pi d}$, or $\pi = \frac{2l}{pd}$.
- We can estimate p by dropping N needles and forming the estimate $\hat{p}^{(N)} = \#\{\text{needle } i \text{ hit}\}/N$
- We can then use this to estimate π: $\hat{\pi} = \frac{2l}{\hat{p}^{(N)} d}$.



Joint PDFs of multiple RVs

The plot was created with the following Mathematica code:

```
(* Buffon's needle *)
Clear[phat, πhat];
Off[Power::infy];
phat[l_, d_, Nn_] := Accumulate[Table[If[RandomReal[{0, d / 2}] ≤ l / 2 Sin[RandomReal[{0, π / 2}]], 1, 0], Range[Nn]]] / Range[Nn]
πhat[l_, d_, Nn_] := 2 l / (phat[l, d, Nn] d);
Clear[l, d, Nn, pl];
l = 1; d = 2; Nn = 1000;
pl = ListPlot[πhat[l, d, Nn], GridLines → {{}, {π}}, PlotRange → {{0, Nn}, {0, 2 π}}]
Export[NotebookDirectory[] <> "Fig-Buffon3.pdf", pl];
```

Outline

Joint CDFs of multiple RVs

Recall. Given a PDF f_X for a continuous RV X we obtain the CDF F_X via

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t.$$

This is more easily understood in analogy with the case of a PMF p_Y for a discrete RV Y, where the CDF F_Y is

$$F_Y(y) = \sum_{x \leq y} p_Y(x).$$

For discrete RVs (X, Y) with joint PMF $p_{X,Y}(x, y)$, the joint CDF is

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \sum_{(u,v):u \leq x, v \leq y} p_{X,Y}(u, v).$$

The PDF is obtainable from the CDF via $f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x)$.

Joint CDFs of multiple RVs

The joint CDF of two continuous RVs (X, Y) is

$$F_{X,Y}(x,y) = \mathbb{P}(X \le x, Y \le y).$$

The joint CDF can be found from the joint PDF by integration:

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s,t) \mathrm{d}t \mathrm{d}s.$$

The joint PDF can be found from the joint CDF by differentiation:

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

Joint CDFs of multiple RVs

Example. Suppose (X, Y) have joint CDF

$$F_{X,Y}(x,y) = (1-\mathrm{e}^{-\lambda x})(1-\mathrm{e}^{-\mu y}), \ (x,y) \in \mathbb{R}_+^2.$$

Find the joint PDF:

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{\partial}{\partial x}\frac{\partial}{\partial y} F_{X,Y}(x,y) \\ &= \frac{\partial}{\partial x}\frac{\partial}{\partial y}(1-\mathrm{e}^{-\lambda x})(1-\mathrm{e}^{-\mu y}) \\ &= \frac{\partial}{\partial x}(1-\mathrm{e}^{-\lambda x})\frac{\partial}{\partial y}(1-\mathrm{e}^{-\mu y}) \\ &= \lambda \mathrm{e}^{-\lambda x} \mu \mathrm{e}^{-\mu y}. \end{aligned}$$

This example has $X \sim \exp(\lambda)$ and $Y \sim \exp(\mu)$, with (X, Y) independent.

Outline

Expectation

Recall that if X is a continuous RV with PDF $f_X(x)$ and $Y = g(X)$ for some function $g(\cdot)$, then

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) \mathrm{d}x.$$

Now let (X, Y) be a pair of continuous RVs with joint PDF $f_{X,Y}(x, y)$ and $Z = g(X, Y)$ for some function $g(\cdot, \cdot)$, then

$$\mathbb{E}[Z] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{X,Y}(x, y) \mathrm{d}x\mathrm{d}y.$$

Expectation

Example. Let $X \sim \exp(1)$ and $Y \sim \exp(1)$, with (X, Y) independent. Define $Z = \sqrt{XY}$. Find $\mathbb{E}[Z]$.

$$\begin{aligned}
\mathbb{E}[Z] &= \int_0^\infty \int_0^\infty \sqrt{xy}\mathrm{e}^{-x}\mathrm{e}^{-y}\mathrm{d}x\mathrm{d}y \\
&= \int_0^\infty \sqrt{x}\mathrm{e}^{-x}\mathrm{d}x \int_0^\infty \sqrt{y}\mathrm{e}^{-y}\mathrm{d}y \\
&= \frac{\sqrt{\pi}}{2} \times \frac{\sqrt{\pi}}{2} = \frac{\pi}{4}.
\end{aligned}$$

The integral $\int_x \sqrt{x}\mathrm{e}^{-x}\mathrm{d}x$ requires use of $\mathrm{erf}(\cdot)$ function.

Expectation

Linearity continues to hold: if X, Y are continuous RVs with means $\mathbb{E}[X]$ and $\mathbb{E}[Y]$, then

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

Outline

More than two RVs

In perfect analogy with the case of two RVs, we have for three RVs (X, Y, Z):

$$\mathbb{P}((X, Y, Z) \in B) = \int_{(x,y,z)\in B} f_{X,Y,Z}(x, y, z)\mathrm{d}x\mathrm{d}y\mathrm{d}z, \ \forall B \subset \mathbb{R}^3.$$

We can marginalize Z by integrating over it to find the joint for (X, Y):

$$f_{X,Y}(x, y) = \int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z)\mathrm{d}z,$$

and we can marginalize (Y, Z) by integrating over them to find the marginal for X:

$$f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z)\mathrm{d}y\mathrm{d}z.$$

More than two RVs

Naturally the expecation for $W = g(X, Y, Z)$ is:

$$\mathbb{E}[W] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y, z) f_{X,Y,Z}(x, y, z)\mathrm{d}x\mathrm{d}y\mathrm{d}z.$$

and naturally

$$\mathbb{E}[aX + bY + cZ] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c\mathbb{E}[Z].$$

More generally, for RVs $(X_1, \ldots, X_n)$ and scalars $(a_1, \ldots, a_n)$ we have:

$$\mathbb{E}[a_1 X_1 + \cdots a_n X_n] = \mathbb{E}\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i \mathbb{E}[X_i] = a_1\mathbb{E}[X_1] + \cdots + a_n\mathbb{E}[X_n].$$

Outline

Outline

Conditioning an RV on an event

Given an event A with $\mathbb{P}(A) > 0$ the conditional PDF $f_{X|A}(x)$ is defined as the function for which:

$$\mathbb{P}(X \in B|A) = \int_B f_{X|A}(x)\mathrm{d}x, \ \forall B \subset \mathbb{R}.$$

Again, by choosing $B = \mathbb{R}$ we require normalization:

$$1 = \int_{-\infty}^{\infty} f_{X|A}(x)\mathrm{d}x.$$

For events of the form $\{X \in A\}$ we find

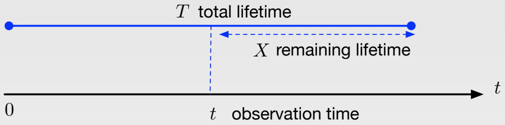
$$\mathbb{P}(X \in B|X \in A) = \frac{\mathbb{P}(X \in B, X \in A)}{\mathbb{P}(X \in A)} = \frac{1}{\mathbb{P}(X \in A)} \int_{A \cap B} f_X(x)\mathrm{d}x$$

which means the conditional PDF is

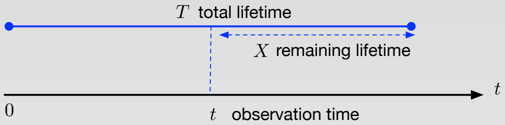
$$f_{X|\{X \in A\}}(x) = \begin{cases} \frac{f_X(x)}{\mathbb{P}(X \in A)}, & x \in A \\ 0, & \text{else} \end{cases}$$

Conditioning an RV on an event

Example. (Exponential RV is memoryless.) Suppose the lifetime of a lightbulb T is an exponential RV with parameter λ, i.e., $T \sim \mathrm{Exp}(\lambda)$. Given $T > t$, find the distribution for the additional lifetime X of the lightbulb.



Conditioning an RV on an event



Let $A = \{T > t\}$. Then:

$$\begin{aligned}
\mathbb{P}(X > x|A) &= \mathbb{P}(T > t + x|T > t) \\
&= \frac{\mathbb{P}(T > t + x, T > t)}{\mathbb{P}(T > t)} \\
&= \frac{\mathbb{P}(T > t + x)}{\mathbb{P}(T > t)} = \frac{\mathrm{e}^{-\lambda(t+x)}}{\mathrm{e}^{-\lambda t}} = \mathrm{e}^{-\lambda x}.
\end{aligned}$$

In other words, $\mathbb{P}(X > x|A) = \mathbb{P}(X > x)$, i.e., the additional lifetime of the lightbulb is independent of the past lifetime. This is the memorylessness property of the exponential distribution.