CS145: Probability & Computing

Lecture 9: Continuous Probability Densities, Gaussian Distributions

Figure credits:
*Bertsekas & Tsitsiklis, **Introduction to Probability**, 2008*
*Pitman, **Probability**, 1999*

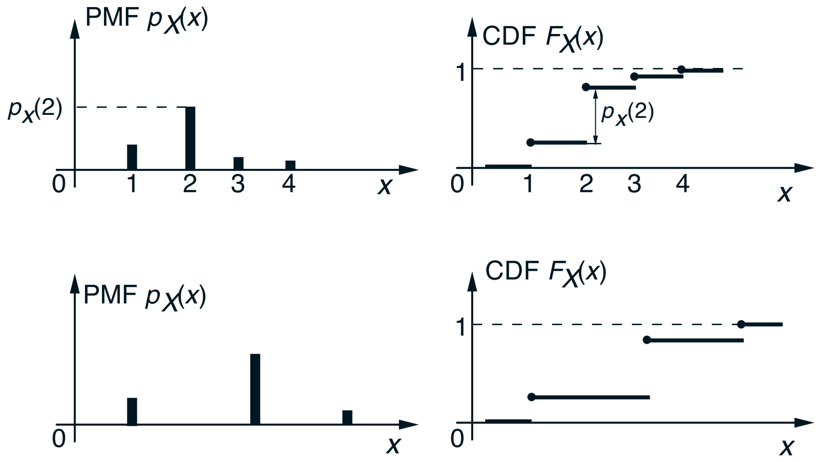
CS145: Lecture 9 Outline

- Continuous Random Variables & Probability Densities
- Gaussian (Normal) Distributions

Cumulative Distribution Function (CDF)

- Recall *probability mass function* (PMF):

$$p_X(x) = P(X = x)$$



- The *cumulative distribution function* (CDF) is the cumulative sum of the PMF:

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

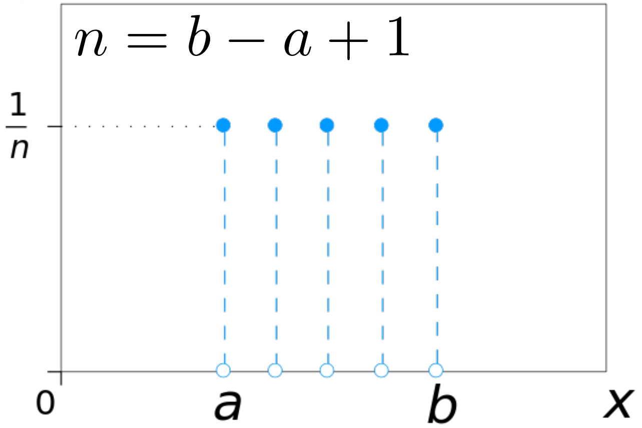
- The CDF equals 0 below the range of X, 1 above the range of X, and is *monotonically increasing*: $$F_X(x_2) \geq F_X(x_1) \text{ if } x_2 > x_1.$$

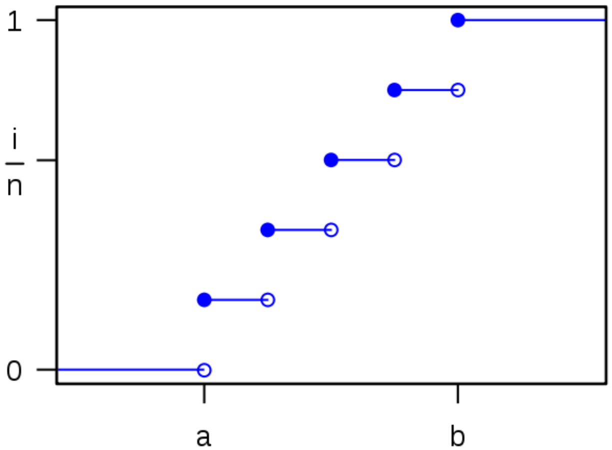
- The CDF allows quick computation of the probability of *intervals*:

$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$$

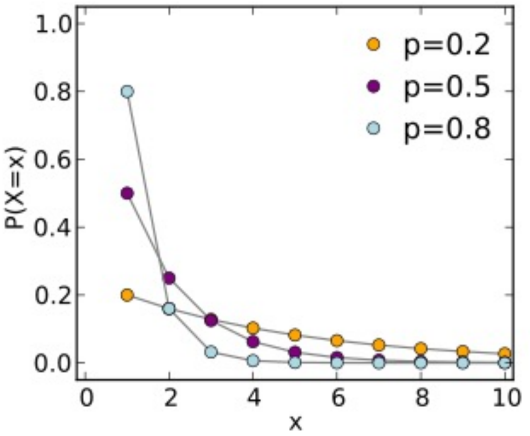
Examples of Discrete PMFs & CDFs

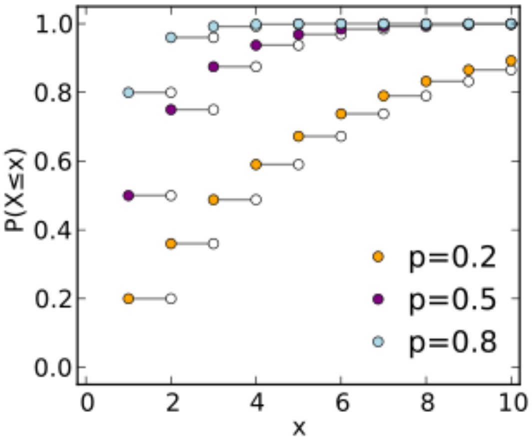
Uniform



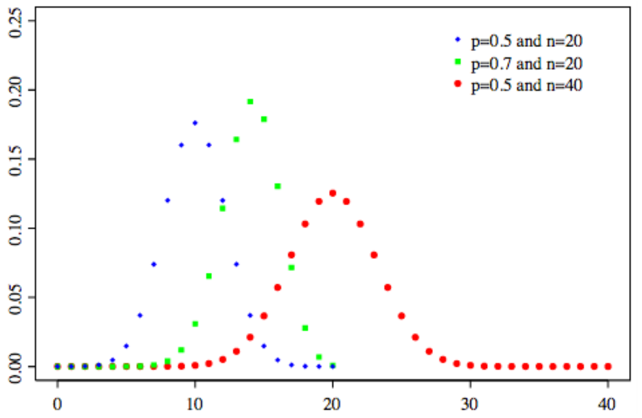


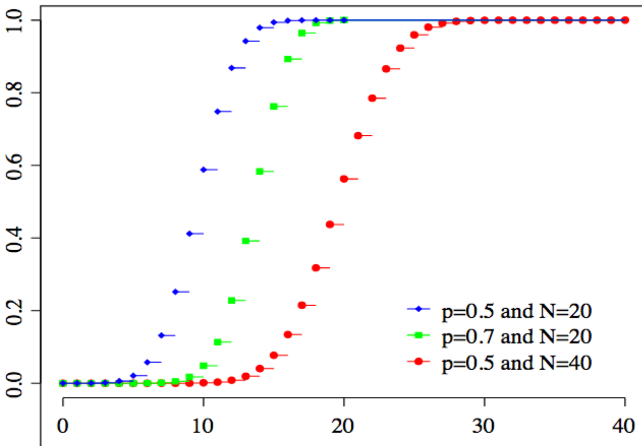
Geometric



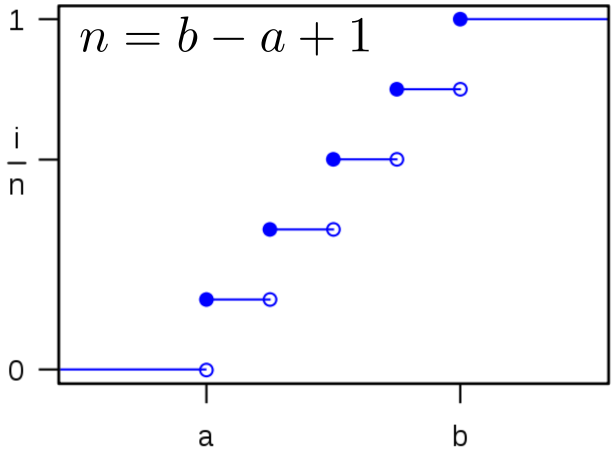


Binomial





Continuous Random Variables

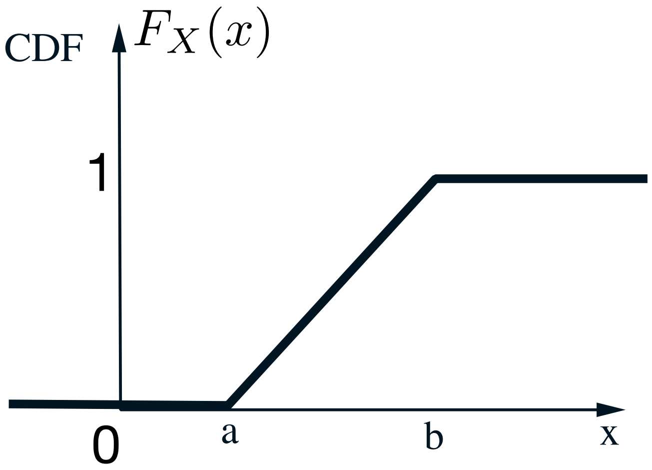


- For any discrete random variable, the CDF is *discontinuous and piecewise constant*
- If the CDF is *continuous**, we have a continuous random variable:

$$0 \leq F_X(x) \leq 1$$

$$F_X(x_2) \geq F_X(x_1) \text{ if } x_2 > x_1.$$

$$\lim_{x \to -\infty} F_X(x) = 0 \qquad \lim_{x \to +\infty} F_X(x) = 1$$

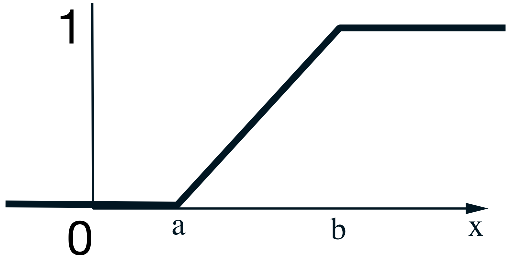


- The probability that continuous random variable *X* lies in the interval *(x_1,x_2]* is then

$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$$

Continuous CDFs Define Probability Laws

Continuous random variables satisfy the *axioms of probability*.



Non-negativity:

$$P(x_1 < X \leq x_2) \geq 0 \text{ for any } x_1, x_2.$$

Normalization:

$$P(-\infty < X < +\infty) = F_X(+\infty) - F_X(-\infty) = 1 - 0 = 1.$$

Countable Additivity: If $x_1 < x_2 < x_3$,

$$P(x_1 < X \leq x_3) = P(x_1 < X \leq x_2) + P(x_2 < X \leq x_3)$$

$$F_X(x_3) - F_X(x_1) = (F_X(x_2) - F_X(x_1)) + (F_X(x_3) - F_X(x_2))$$

- The probability that continuous random variable *X* lies in interval $(x_1, x_2]$ is

$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$$

Borel Set of Intervals

We can assume sample space $\Omega = (-\infty, \infty)$.

The CDF assigns a probability to each interval $(-\infty, x]$.

Since $Pr(\Omega) = 1$, it also defines probability to each interval $[x, \infty)$

Additivity defines probability for any interval $[x_1, x_2]$

Countable additivity defines probabilities for any countable union and intersection of intervals.

Probability Density Function (PDF)

- If the CDF is differentiable, its first derivative is called the *probability density function (PDF)*:

$$f_X(x) = \frac{dF_X(x)}{dx} = F'_X(x)$$

- By the *fundamental theorem of calculus*:

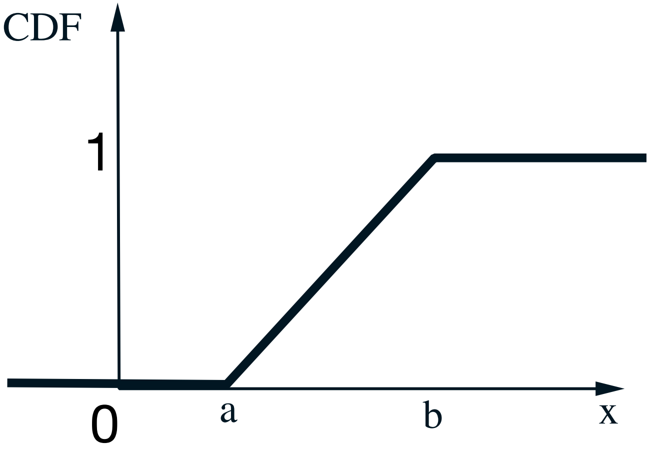
$$\int_{x_1}^{x_2} f_X(x)\, dx = F_X(x_2) - F_X(x_1) = P(x_1 < X \leq x_2)$$

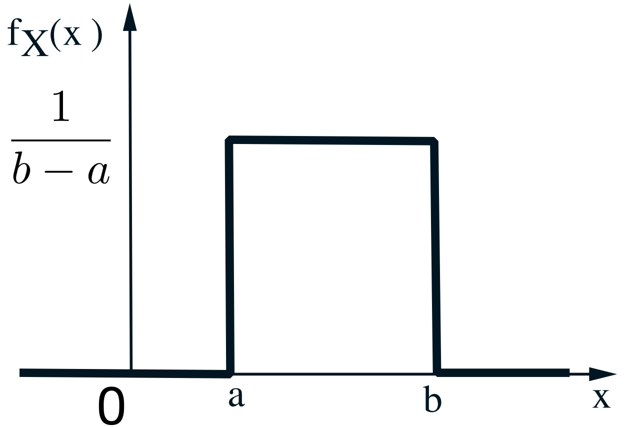
- For any valid PDF:

$$f_X(x) \geq 0$$

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

$$0 \leq \int_{x_1}^{x_2} f_X(x)\, dx \leq 1$$





Continuous Uniform Probability Densities

➢ Relationship between CDF and PDF:

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx \qquad f_X(x) = \frac{dF_X(x)}{dx}$$

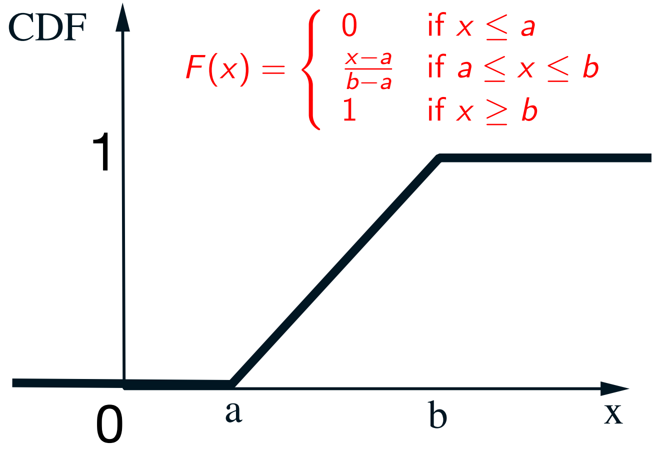
➢ For a *continuous uniform random variable*, an interval's probability is proportional to length:

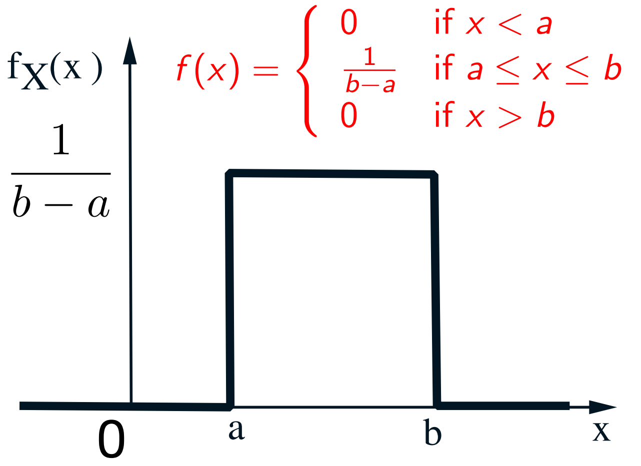
$$\text{If } a \le x_1 < x_2 \le b,$$

$$P(x_1 < X \le x_2) = \frac{x_2 - x_1}{b - a}$$

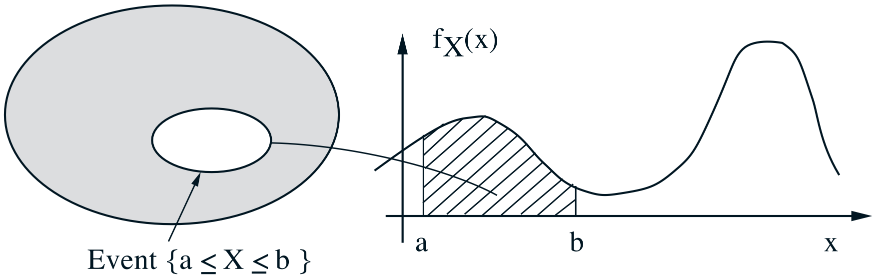
➢ Note that it is possible that $f_X(x) > 1$

$$\text{If } a = 0 \text{ and } b = 0.1, \text{ then } \frac{1}{b-a} = 10.$$

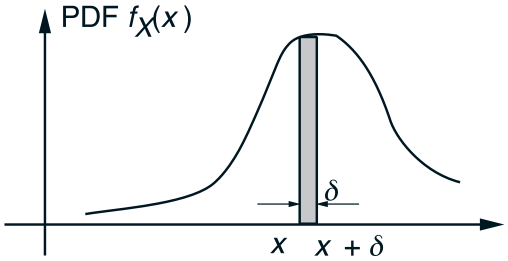




General Continuous Random Variables



$$f_X(x) \geq 0$$



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

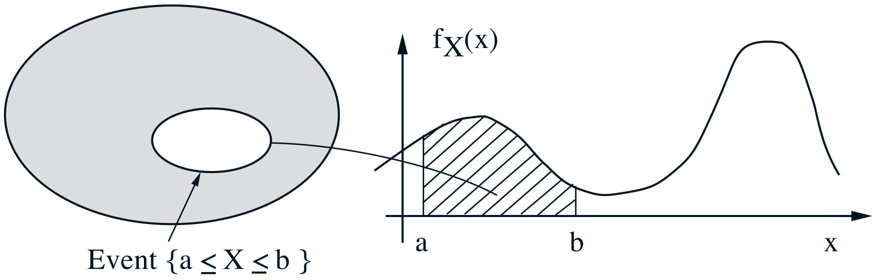
$$\mathbf{P}(x \leq X \leq x+\delta) = \int_x^{x+\delta} f_X(s)\,ds \approx f_X(x) \cdot \delta$$

Observation: For a continuous random variable, the probability of observing X=x for any particular real number x equals zero:

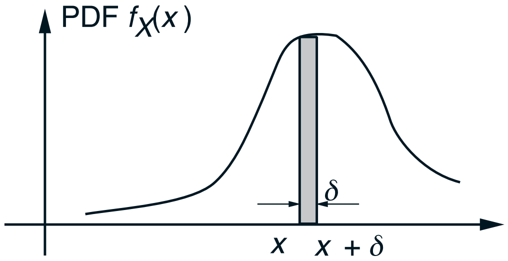
$$P(X=x) = \lim_{\delta \to 0} P(x-\delta < X \leq x) = \lim_{\delta \to 0} \int_{x-\delta}^{x} f_X(s)\,ds = 0$$

As floating point precision increases, probability of any particular number decreases.

General Continuous Random Variables



$$f_X(x) \geq 0$$



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

$$\mathbf{P}(x \leq X \leq x + \delta) = \int_x^{x+\delta} f_X(s)\,ds \approx f_X(x) \cdot \delta$$

Observation: A PDF may take on arbitrarily large positive values:

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

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^1 \frac{1}{2\sqrt{x}}\,dx = \sqrt{x}\Big|_0^1 = 1.$$

Expectations of Continuous Variables

- The *expectation* or *expected value* of a continuous random variable is:

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

- The expected value of a function of a continuous random variable:

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

- The variance of a continuous random variable:

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = E[(X - E[X])^2] = \int_{-\infty}^{+\infty} (x - E[X])^2 f_X(x)\, dx$$

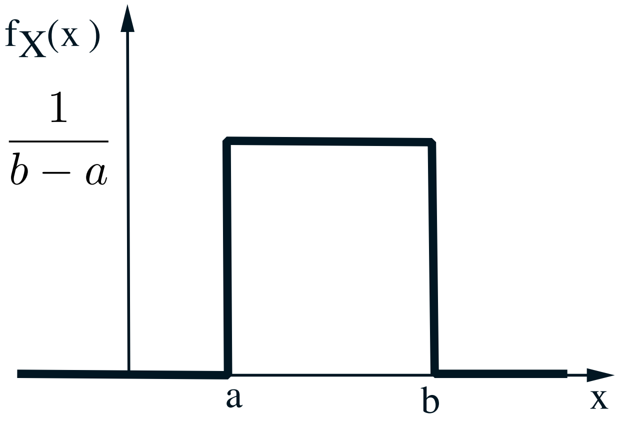
- **Intuition:** Create a discrete variable by quantizing X and compute discrete expectation. As number of discrete values grows, sum approaches integral.

Moments of Uniform Distribution

The expectation of X is

$$\mathbf{E}[X] = \int_a^b \frac{x}{b-a}dx = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2},$$

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



and the second moment is

$$\mathbf{E}[X^2] = \int_a^b \frac{x^2}{b-a}dx = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3}.$$

The variance is computed by

$$\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 = \frac{b^2+ab+a^2}{3} - \frac{(b+a)^2}{4}$$

$$= \frac{(b-a)^2}{12}.$$

Conditioning a Uniform Distribution

Lemma

Let X be a uniform random variable on $[a, b]$. Then for $c \leq d$

$$\Pr(X \leq c \,|\, X \leq d) = \frac{c-a}{d-a}.$$

That is, conditioned on the fact that $X \leq d$, X is uniform on $[a, d]$.

Proof.

$$\begin{aligned}
\Pr(X \leq c \,|\, X \leq d) &= \frac{\Pr((X \leq c) \cap (X \leq d))}{\Pr(X \leq d)} \\
&= \frac{\Pr(X \leq c)}{\Pr(X \leq d)} \\
&= \frac{c-a}{d-a}.
\end{aligned}$$

Exponential Distribution

Definition

The exponential distribution with parameter θ:

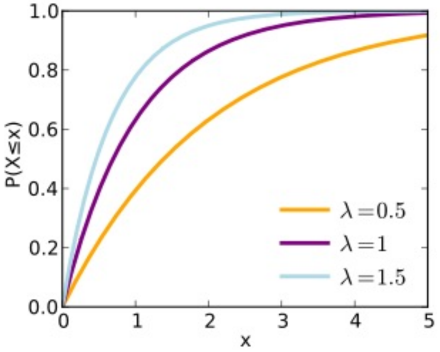
$$F(x) = \begin{cases} 1 - \mathrm{e}^{-\theta x} & \text{for } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

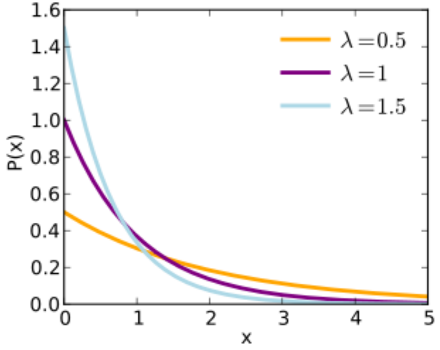
$$f(x) = \theta \mathrm{e}^{-\theta x}, \text{ for } x \geq 0.$$

$$E[X] = \int_0^{\infty} t\theta \mathrm{e}^{-\theta t} dt = \frac{1}{\theta}.$$

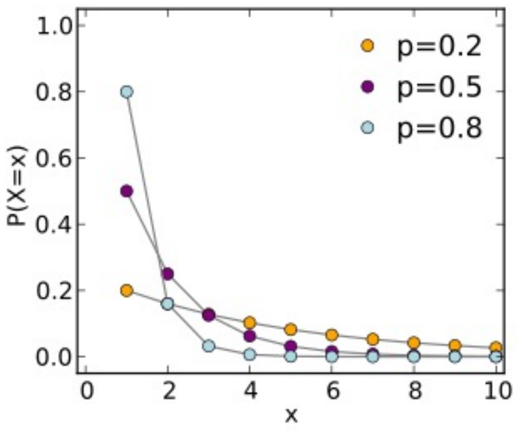
$$E[X^2] = \int_0^{\infty} t^2\theta \mathrm{e}^{-\theta t} dt = \frac{2}{\theta^2}.$$

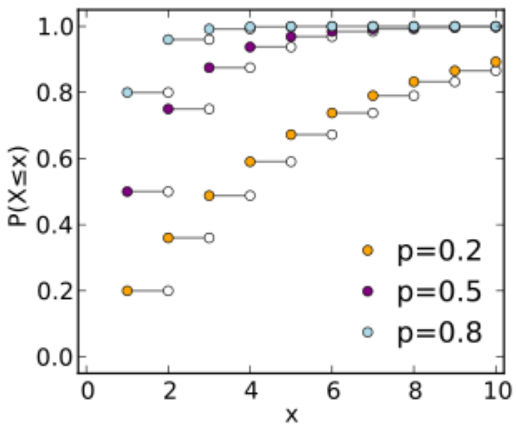
$$\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2 = \frac{1}{\theta^2}.$$

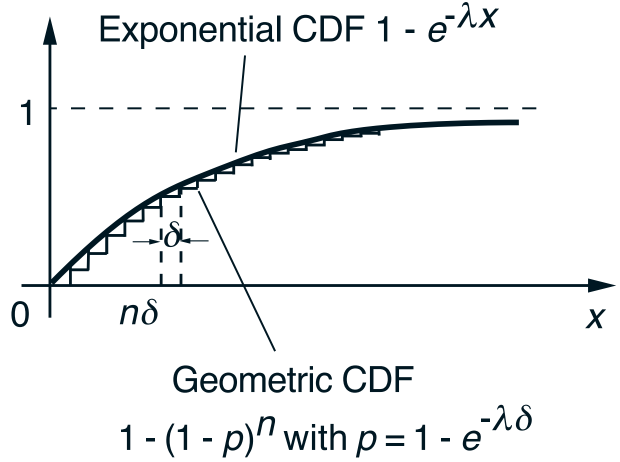




From Geometric to Exponential







- With the matched exponential and geometric parameters given above:

$$F^{\exp}(n\delta) = F^{\mathrm{geo}}(n), \qquad n = 1, 2, \ldots,$$

- Interpretation: If we very quickly toss a coin (every $\delta \ll 1$ seconds) toss a coin with a very small probability of coming up heads, the distribution of the time until the first head is approximately exponential

Exponential Distributions are Memoryless

Lemma

For an exponential random variable with parameter θ,

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

Proof.

$$\begin{aligned}
\Pr(X > s + t \mid X > t) &= \frac{\Pr(X > s + t)}{\Pr(X > t)} \\
&= \frac{1 - \Pr(X \leq s + t)}{1 - \Pr(X \leq t)} \\
&= \frac{\mathrm{e}^{-\theta(s+t)}}{\mathrm{e}^{-\theta t}} \\
&= \mathrm{e}^{-\theta s} = \Pr(X > s).
\end{aligned}$$

"Theory"

Advanced topic not covered in homeworks or exams!

Probability Space $(\Omega, \mathcal{F}, P)$

- Ω - set of all possible outcomes.
- $\mathcal{F}$ set of "allowable" (measurable) events. Must be a σ-field.
  - $\emptyset \in \mathcal{F}$
  - Closed under complements: if $A \in \mathcal{F}$ then $\bar{A} = \Omega \setminus A \in \mathcal{F}$
  - Closed under countable unions (and intersections).
  - Example: Borel Set - set of all open intervals in $\mathcal{R}$.
- P - probability function $P : \mathcal{F} \to [0, 1]$.
  - $P(\Omega) = 1$
  - P is countably additive: for any countable collection of disjoint sets $A_i \in \mathcal{F}$
    $$P(\cup_i A_i) = \sum_i P(A_i).$$

CS145: Lecture 9 Outline

- Continuous Random Variables & Probability Densities
- Gaussian (Normal) Distributions

Gaussian (Normal) Distributions

The density function of the **Normal distribution** $N(\mu, \sigma^2)$ is:

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

The distribution function:

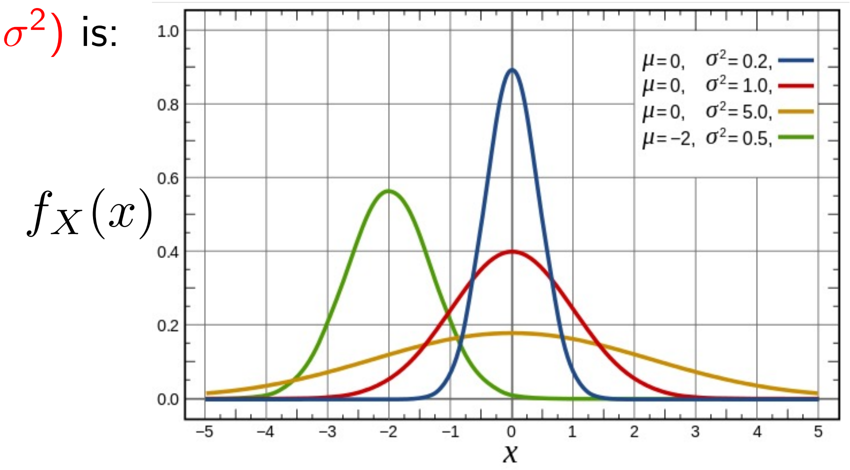
$$F_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

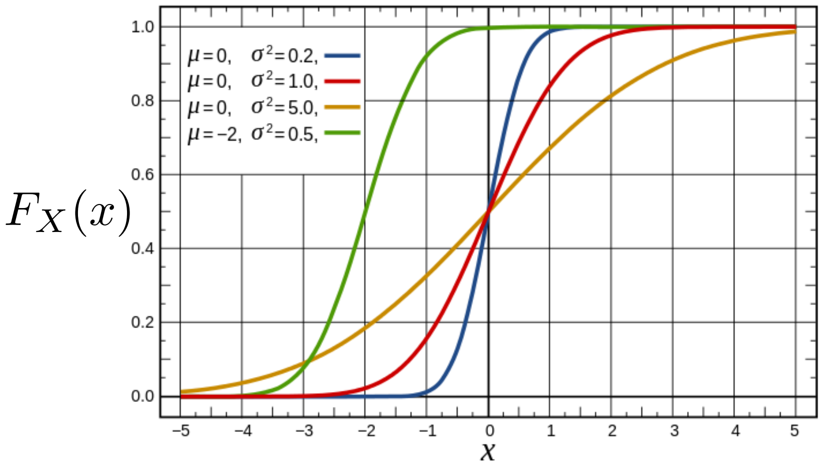
Properties:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1.$$

$$E(X) = \mu \qquad\qquad Var(X) = \sigma^2$$

The integral has no closed form.





Some History (≈ 1700)

The de Moivre-Laplace theorem:

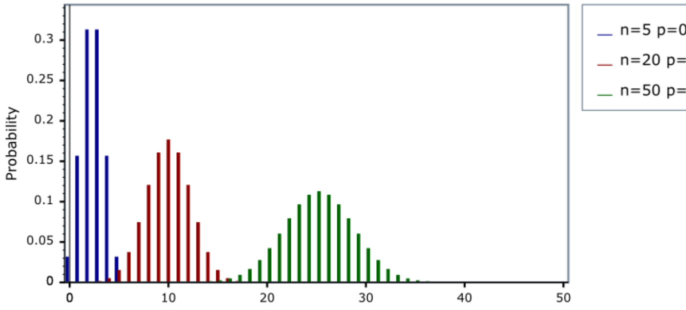
Theorem

For $k = np \pm O(\sqrt{npq})$, $q = 1 - p$:

$$\lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}$$

Setting $\mu = np$ and $\sigma^2 = npq$

$$\lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(k-\mu)^2}{2\sigma^2}}$$



Note: that's discrete probability, not density!

Why the Normal Distribution?

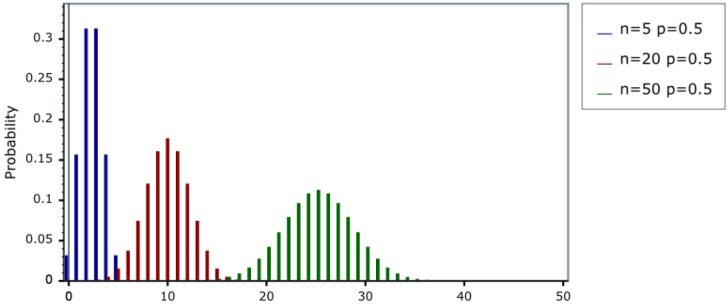
Empirical observation: Many random phenomena follow (at least approximately) Normal distribution.

- Height, weight, income,....
- The velocity of molecule in gas (Brownian Motion)
- Measurement error, noise...
- ....

The Central Limit Theorem:
"The distribution of the average of large number of independent random variable converges to the Normal distribution".

Binomial Distribution:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n$$



Normal Binomial Approximation

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n$$

Let $t = \frac{k-np}{\sqrt{nqp}}$, and "pretend that k is continuous", then $dk/dt = \sqrt{nqp}$.

$$\lim_{n\to\infty} Pr(np - a\sqrt{npq} \le k \le np + b\sqrt{npq})$$

$$\approx \sum_{k=np-a\sqrt{npq}}^{np+b\sqrt{npq}} \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}} \approx \int_{k=np-a\sqrt{npq}}^{np+b\sqrt{npq}} \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}} dk$$

$$\approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt$$

The Normal is a Proper Distribution

Non-negative: since the density $\frac{1}{\sqrt{2\pi\sigma^2}}\ e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$ is non-negative so is the CDF.

Next we need to show that

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

Let $y = (x-\mu)/\sigma$, then we need to show that $I = \int_{-\infty}^{\infty} e^{-y^2/2} dy = \sqrt{2\pi}$.

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(s^2+z^2)/2} ds dz$$

.

Set $s = r\cos\theta$, $z = r\sin\theta$,

$$I^2 = \int_0^{\infty}\int_0^{2\pi} e^{-r^2/2} r\ d\theta\ dr = 2\pi$$

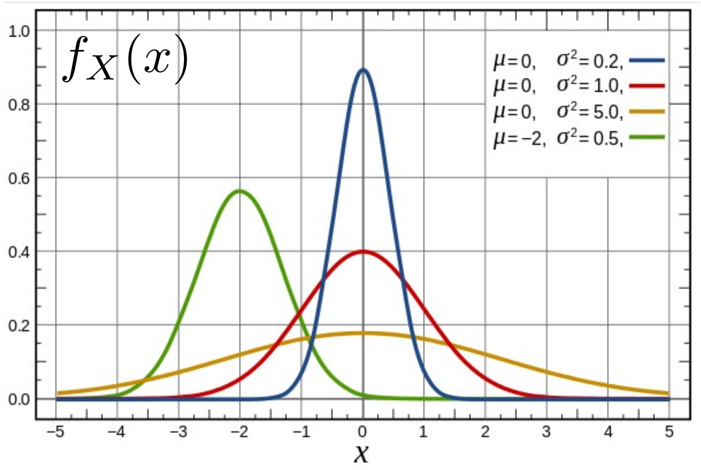
Scaling a Gaussian Variable

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

$$E[X] = \mu$$

$$\text{Var}[X] = E[(X-\mu)^2] = \sigma^2$$

$$\sqrt{\text{Var}[X]} = \sigma \text{ is the standard deviation}$$



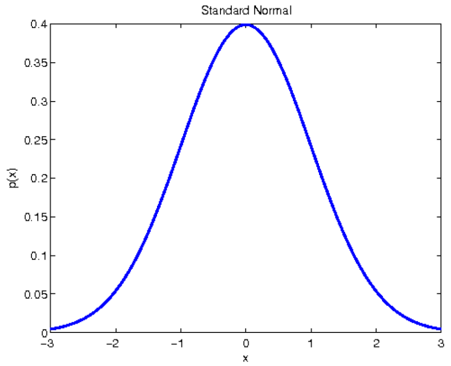
Any linear transformation of a Gaussian variable is Gaussian!

$$Y = aX + b \qquad\qquad f_Y(y) = \frac{1}{\sqrt{2\pi\bar{\sigma}^2}} e^{-\frac{1}{2}\left(\frac{y-\bar{\mu}}{\bar{\sigma}}\right)^2}$$

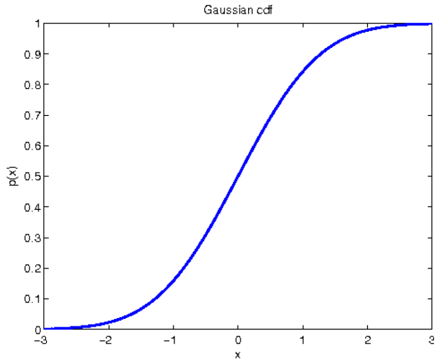
- Mean and variance of linear functions: $\bar{\mu} = a\mu + b, \qquad \bar{\sigma} = |a|\sigma$
- Proof that PDF is Gaussian will come later …

Standard Normal Random Variables

- If $X \sim N(\mu, \sigma^2)$ then for any constants a and b the random variable $aX + b$ is distributed $N(a\mu + b, a^2\sigma^2)$.
- If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma}$ is distribution $N(0, 1)$
- $N(0, 1)$ is the standard Normal distribution.



$$Pr(Z \leq z) = \Phi_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$$



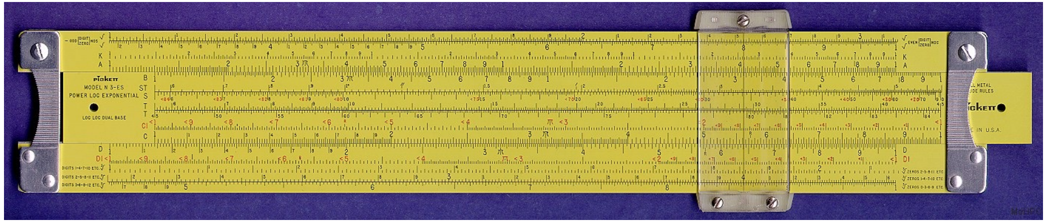
$$\phi_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Classic Computation of Normal CDF

- If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma}$
- $$Pr(X \leq x) = Pr(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}) = \Phi(\frac{x-\mu}{\sigma})$$
- The standard Normal random variable is symmetric around 0.
- For $\frac{x-\mu}{\sigma} < 0$,
$$\Phi(\frac{x-\mu}{\sigma}) = 1 - \Phi(-\frac{x-\mu}{\sigma})$$
- With a table of $\Phi(Z)$ for $Z > 0$ we can compute $F(x)$ for any Normal random variable

Probability Content
from -oo to Z

Z	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
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990



Modern Computation of Normal CDF

normcdf

R2015b

Normal cumulative distribution function

collapse all in page

Syntax

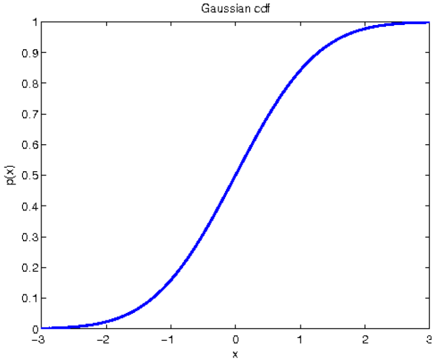
```
p = normcdf(x)
p = normcdf(x,mu,sigma)
[p,plo,pup] = normcdf(x,mu,sigma,pcov,alpha)
[p,plo,pup] = normcdf(___,'upper')
```

Description

`p = normcdf(x)` returns the standard normal cdf at each value in `x`. The standard normal distribution has parameters `mu = 0` and `sigma = 1`. `x` can be a vector, matrix, or multidimensional array.

`p = normcdf(x,mu,sigma)` returns the normal cdf at each value in `x` using the specified values for the mean `mu` and standard deviation `sigma`. `x`, `mu`, and `sigma` can be vectors, matrices, or multidimensional arrays that all have the same size. A scalar input is expanded to a constant array with the same dimensions as the other inputs. The parameters in `sigma` must be positive.

$$p = \Phi(x)$$



$$x = \Phi^{-1}(p)$$

norminv

R2015b

Normal inverse cumulative distribution function

collapse all in page

Syntax

```
X = norminv(P,mu,sigma)
[X,XLO,XUP] = norminv(P,mu,sigma,pcov,alpha)
```

Description

`X = norminv(P,mu,sigma)` computes the inverse of the normal cdf using the corresponding mean `mu` and standard deviation `sigma` at the corresponding probabilities in `P`. `P`, `mu`, and `sigma` can be vectors, matrices, or multidimensional arrays that all have the same size. A scalar input is expanded to a constant array with the same dimensions as the other inputs. The parameters in `sigma` must be positive, and the values in `P` must lie in the interval [0 1].

Moment Generating Function

Definition

For any random variable X the **Moment Generating Function** of X is

$$M_X(t) = E[e^{tX}].$$

Theorem

If $M_X(t)$ exists in some interval $(-\delta, \delta)$, then for any $n \geq 1$,

$$\frac{d^k M_X(t)}{dt} \Big|_{t=0} = E[X^k].$$

Proof

Theorem

If $M_X(t)$ *exists in some interval* $(-\delta, \delta)$, *then for any* $n \geq 1$,

$$\frac{d^k M_X(t)}{dt} \mid_{t=0} = E[X^k].$$

Proof.

$$\frac{d^k M_X(t)}{dt} = \frac{d^k E[e^{tX}]}{dt} = E[\frac{d^k e^{tX}}{dt}] = E[X^k e^{tX}]$$

$$\frac{d^k M_X(t)}{dt} \mid_{t=0} = E[X^k e^{tX}] \mid_{t=0} = E[X^k]$$

$\square$

Generating Function of a Sum

Theorem

Let X, Y *be independent random variable then*

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

.

Proof.

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX}]E[e^{tY}].$$

$\square$

Examples

Assume $Pr(X = 1) = p$, $Pr(X = 0) = 1 - p$, then

$$M_X(t) = pe^t + (1 - p)$$

$$M'_X(t) \mid_{t=0} = pe^t \mid_{t=0} = p \qquad M''_X(t) \mid_{t=0} = pe^t \mid_{t=0} = p$$

Let $X \sim B(n, p)$. $M_X(t) = (pe^t + (1 - p))^n$

$$M'_X(t) \mid_{t=0} = npe^t(pe^t + (1 - p))^{n-1} \mid_{t=0} = np$$

$$\begin{aligned} & M''_X(t) \mid_{t=0} \\ = \; & [npe^t(pe^t + (1 - p))^{n-1} + n(n - 1)p^2e^{2t}(pe^t + (1 - p))^{n-2}] \mid_{t=0} \\ = \; & np + n(n - 1)p^2 \end{aligned}$$

Back to the Normal Distribution

We first compute the moment generating function of $x \sim N(0,1)$.

$$
\begin{aligned}
M_x(t) &= E[e^{tx}] \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2} + \frac{t^2}{2}} dx \\
&= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}} dx \\
&= e^{t^2/2}
\end{aligned}
$$

Moments of the Normal Distribution

Assume now that $x \sim N(\mu, \sigma^2)$. Let $z = \frac{x-\mu}{\sigma}$, then $\sigma dz = dx$.

$$
\begin{aligned}
M_x(t) &= E[e^{tx}] \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{t\sigma z + t\mu} e^{-\frac{z^2}{2}} \sigma dz \\
&= \frac{1}{\sqrt{2\pi}} e^{\mu t} \int_{-\infty}^{\infty} e^{-\frac{(z-\sigma t)^2}{2} + \frac{(\sigma t)^2}{2}} dz \\
&= e^{\frac{t^2\sigma^2}{2} + \mu t}
\end{aligned}
$$

Expectation and Variance of N(μ,σ)

$$M_x'(t) = (\mu + t\sigma^2)e^{\frac{t^2\sigma^2}{2}+\mu t}$$

$$M_X''(T) = (\mu + t\sigma^2)^2 e^{\frac{t^2\sigma^2}{2}+\mu t} + \sigma^2 e^{\frac{t^2\sigma^2}{2}+\mu t}$$

$$E[x] = M'(0) = \mu$$

$$E[x^2] = M''(0) = \mu^2 + \sigma^2$$

.

$$Var[x] = E[x^2] - (E[x])^2 = \sigma^2$$

M.G.F Defines a Distribution

Theorem

Let X *and* Y *be two random variables. If*

$$M_X(t) = M_Y(t)$$

for all $t \in (-\delta, \delta)$ *for some* $\delta > 0$, *then* X *and* Y *have the same distribution.*

Sum of Normal Random Variables

Theorem

Let X, Y be independent random variables with $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ then

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Proof.

$$M_{X+Y}(t) = e^{\frac{t^2\sigma_1^2}{2}+\mu_1 t} e^{\frac{t^2\sigma_2^2}{2}+\mu_2 t} = e^{\frac{t^2(\sigma_1^2+\sigma_2^2)}{2}+(\mu_1+\mu_2)t}.$$

□