Lecture 2

Random variables: discrete and continuous

Random variables: discrete

Probability theory is concerned with situations in which the outcomes occur randomly. Generically, such situations are called *experiments*, and the set of all possible outcomes is the **sample space** corresponding to an experiment.

- Tossing a coin 3 times:

$$\Omega = \{hhh, hht, htt, hth, ttt, tth, thh, tht\}$$

- A random variable is a function from sample space to real numbers:
  - Total # (number) of heads
  - Total # of tails
  - # of heads minus # of tails
  - …

- Discrete random variables can take only **finite or countably infinite** # of values
  - Toss a coil until a head turns up, # of tosses is countably infinite
  - A countably infinite set: one-to-one correspondence with the integers

Random variables: discrete

- Tossing a coin 3 times:

$$\Omega = \{hhh, hht, htt, hth, ttt, tth, thh, tht\}$$

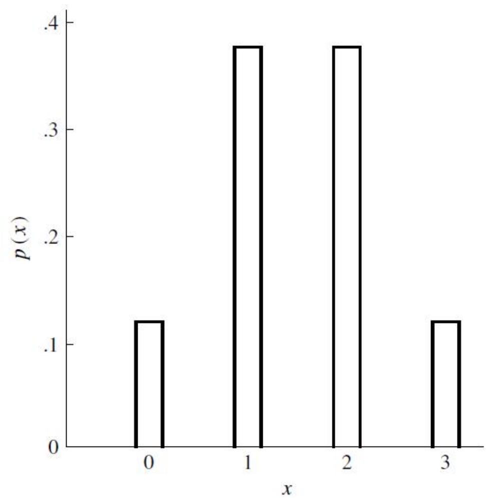
X=total # of heads

$$P(X=0) = \tfrac{1}{8}$$
$$P(X=1) = \tfrac{3}{8}$$
$$P(X=2) = \tfrac{3}{8}$$
$$P(X=3) = \tfrac{1}{8}$$

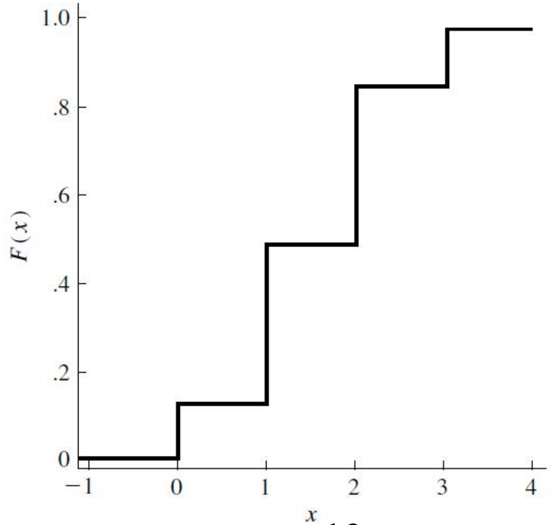
- Probability mass function or frequency function
- Cumulative distribution function (**cdf**, non-decreasing)

$$F(x) = P(X \le x), \qquad -\infty < x < \infty$$

$$\lim_{x \to -\infty} F(x) = 0 \qquad \text{and} \qquad \lim_{x \to \infty} F(x) = 1.$$



frequency function



cdf

Bernoulli Random Variables

An experiment is either a success (1, probability p) or a failure (0):

$$p(1) = p$$
$$p(0) = 1 - p$$
$$p(x) = 0, \qquad \text{if } x \neq 0 \text{ and } x \neq 1$$

- An alternative representation

$$p(x) = \begin{cases} p^x(1-p)^{1-x}, & \text{if } x = 0 \text{ or } x = 1 \\ 0, & \text{otherwise} \end{cases}$$

- *Indicator random variable*

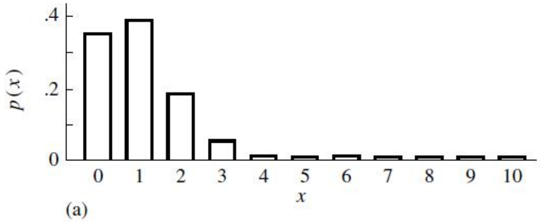
- Examples:
  - Product quality assessment: pass/fail
  - A new-born: female/male
  - Computer commands in binary form: 1/0
  - Schrödinger's cats… ??

The Binomial Distribution

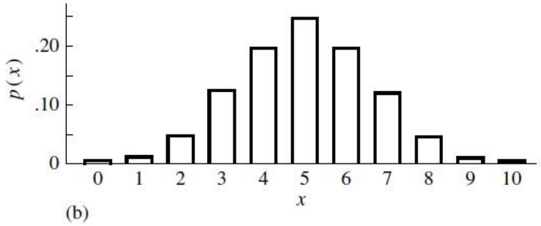
Perform n (fixed) success/failure experiments. X=how many successes.

Example: 10 coin tosses, how many led to a "head"?

- For a given X=k, any particular sequence of k successes occurs with probability $p^k(1-p)^{n-k}$
- The total # of such sequences is $\binom{n}{k}$
- Finally $p(k) = \binom{n}{k} p^k (1-p)^{n-k}$



$n = 10$ and $p = 0.1$



$n = 10$ and $p = 0.5$

The Binomial Distribution

Example. If a single bit (0 or 1) is transmitted over a noisy communications channel, it has probability p of being incorrectly transmitted.

To improve the reliability of the transmission, the bit is transmitted n times, where n is odd. A decoder at the receiving end, called a majority decoder, decides that the correct message is that carried by a majority of the received bits.

Each bit is independently subject to being corrupted with the same probability p. The number of bits that is in error, X, is thus a binomial random variable with n trials and probability p of "success" on each trial (here a "success" is an error).

If n=5, p=0.1, what is the probability that the message is correctly received?

The Binomial Distribution

Example. If a single bit (0 or 1) is transmitted over a noisy communications channel, it has probability p of being incorrectly transmitted.

To improve the reliability of the transmission, the bit is transmitted n times, where n is odd. A decoder at the receiving end, called a majority decoder, decides that the correct message is that carried by a majority of the received bits.

Each bit is independently subject to being corrupted with the same probability p. The number of bits that is in error, X, is thus a binomial random variable with n trials and probability p of "success" on each trial (here a "success" is an error).

If n=5, p=0.1, what is the probability that the message is correctly received?

Yeah!! = 0, 1 or 2 failures,
k failures among n: binomial

$$\sum_{k=0}^{2} \binom{n}{k} p^k (1-p)^{n-k}$$

$$= p^0(1-p)^5 + 5p(1-p)^4 + 10p^2(1-p)^3$$

$$= .9914$$

Much better reliability.

The Geometric Distribution

In independent Bernoulli trials, keep trying until first success.
X=how many trials (including the success).

For X=k, k-1 failures followed by a success. Due to independence,

$$p(k) = P(X = k) = (1 - p)^{k-1} p, \qquad k = 1, 2, 3, \ldots$$

The Negative Binomial Distribution

Keep trying until succeed r times. X=how many trials (including the success).

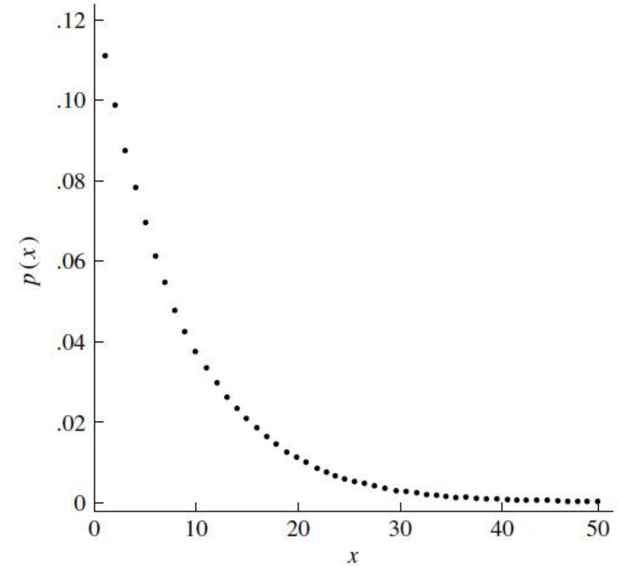
Generalization of geometric distribution. For X=k, r-1 successes assigned to k-1 trails, before the last success.

$$P(X = k) = \binom{k-1}{r-1} p^r (1 - p)^{k-r}$$

Negative binomial random variable=sum of independent geometric variables:
Example: f f f f f s f f s f f f f f f f s f s… (f=failure, s=success)

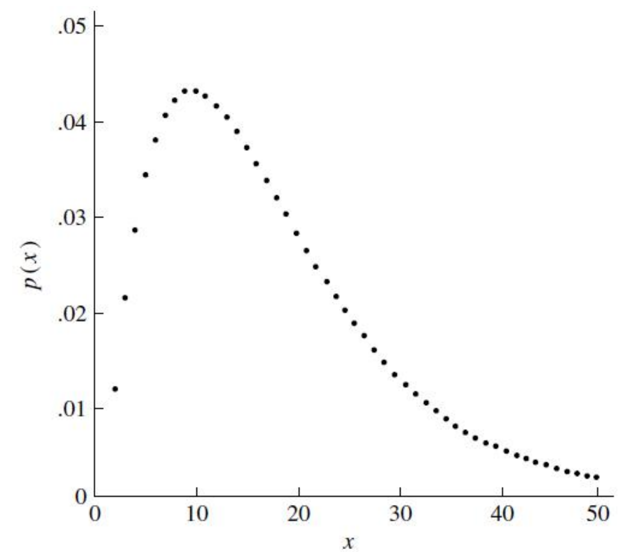
Example A: The probability of winning in a certain state lottery is p=1/9. The distribution of the number of tickets a person must purchase is a geometric random variable:

$$p(k) = P(X = k) = (1 - p)^{k-1} p$$



Example B: He/she wants to win twice! Then the distribution is negative binomial with p=1/9, r=2:

$$p(k) = (k - 1)p^2(1 - p)^{k-2}$$



The Hypergeometric Distribution

Suppose that an urn contains n balls, of which r are black and n-r are white. Let X denote the number of black balls drawn when taking m balls *without replacement*. Then X follows the hypergeometric distribution:

$$P(X = k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$$

When n approaches infinity, let p=r/n, this distribution approximates a binomial distribution with p: Take m balls one by one, it may succeed (black) or fail (white).

The Poisson Distribution

（参见北京大学数学系汪仁官《概率论引论》）
Rutherford和Geiger观察放射性物质发射α粒子时发现，在规定的一段时间内（7.5秒），放射的粒子数 X 服从Poisson分布。为什么？

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}, \qquad k=0,1,2,\ldots$$

把体积V的放射性物质设想分割为n份相同体积的小块$\Delta V=V/n$. 假定：

1. 对于每个小块而言，在7.5s内放出一个粒子的概率p正比于ΔV;
2. 在7.5s内每个小块放出>1个粒子的概率极低，可认为是0;
3. 各个小块是否放出粒子，是相互独立的。

在7.5s内，体积为V的放射物质放出k个粒子这个事件，可近似看作n个独立小块中，恰有k个小块放出粒子，其他不放，所以近似服从二项分布（无限细分，得到精确表达）：

$$p(k)=\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \qquad n\to\infty$$

The Poisson Distribution

$$p(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \qquad n \to \infty$$

记 $np=\lambda$（反映总体而非小块的放射性），

$$p(k) = \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!} \frac{n!}{(n-k)!} \frac{1}{n^k} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k}$$

$$\to \frac{\lambda^k e^{-\lambda}}{k!}$$

$$\frac{\lambda}{n} \to 0$$

$$\frac{n!}{(n-k)!n^k} \to 1$$

$$\left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}$$

$$\left(1-\frac{\lambda}{n}\right)^{-k} \to 1$$

Poisson分布是二项分布当n趋于无穷，$np=\lambda$时的极限分布；

所以，n较大，p较小时，可用Poisson分布来作二项分布的近似计算。

The Poisson Distribution

The # of independent events that happen in a time interval or space volume.

- 电话系统：交换机服务大量客户，客户行为独立，单位时间内交换机被呼叫的次数；
- 保险建模：例如合肥市包河区群众略诡异的事故（淋浴时摔跤，被雷击中……）稀少且独立发生；
- 智能交通：交通不太拥挤，单位时间内通过街道/路口的车辆数;
- CCD接收到的光子数（如X-ray天文学）…

The Poisson Distribution

The # of independent events that happen in a time interval or space volume.

- 电话系统：交换机服务大量客户，客户行为独立，单位时间内交换机被呼叫的次数；
- 保险建模：例如合肥市包河区群众略诡异的事故（淋浴时摔跤，被雷击中……）稀少且独立发生；
- 智能交通：交通不太拥挤，单位时间内通过街道/路口的车辆数;
- CCD接收到的光子数（如X-ray天文学）…

例 2.1.5.2 这个有趣的经典例子来自于 von Bortkiewicz(1898). 记录下 10 个普鲁士骑兵军团在 20 年内被马踢死的士兵数目，共采集到 200 个军团–年的数据. 这些数据和 $\lambda = 0.61$ 的泊松模型概率列示在下表中. 表中第一列给出每年死亡的数目，从 0 到 4. 第二列列示了死亡数目观测到的次数. 因此，例如，在 200 个军团–年的数据中，观测到 1 人死亡的次数是 65. 在表中第三列，将观测到的死亡数目用 200 转化为相对频率. 第四列给出了参数为 $\lambda = 0.61$ 的泊松概率. 在第 8 章和第 9 章，我们讨论如何根据观测频率选择理论概率模型的参数值，以及评价拟合好坏的方法. 目前，我们只是说选择 $\lambda = 0.61$ 与每年平均死亡人数相匹配. ■

每年死亡人数	观 测	相对频率	泊松概率
0	109	0.545	0.543
1	65	0.325	0.331
2	22	0.110	0.101
3	3	0.015	0.021
4	1	0.005	0.003

Poisson and Binomial Distributions

But how large is a large n, how small is a small p?

Example: Two dice are rolled 100 times, count the # of double sixes as X. The distribution of X is binomial with n=100, p=1/36 (why?). Let's compare the two distributions (λ=np=2.78).

Why should we do this? Poisson distribution is easier to parameterize and calculate.

k	Binomial Probability	Poisson Approximation
0	.0596	.0620
1	.1705	.1725
2	.2414	.2397
3	.2255	.2221
4	.1564	.1544
5	.0858	.0858
6	.0389	.0398
7	.0149	.0158
8	.0050	.0055
9	.0015	.0017
10	.0004	.0005
11	.0001	.0001

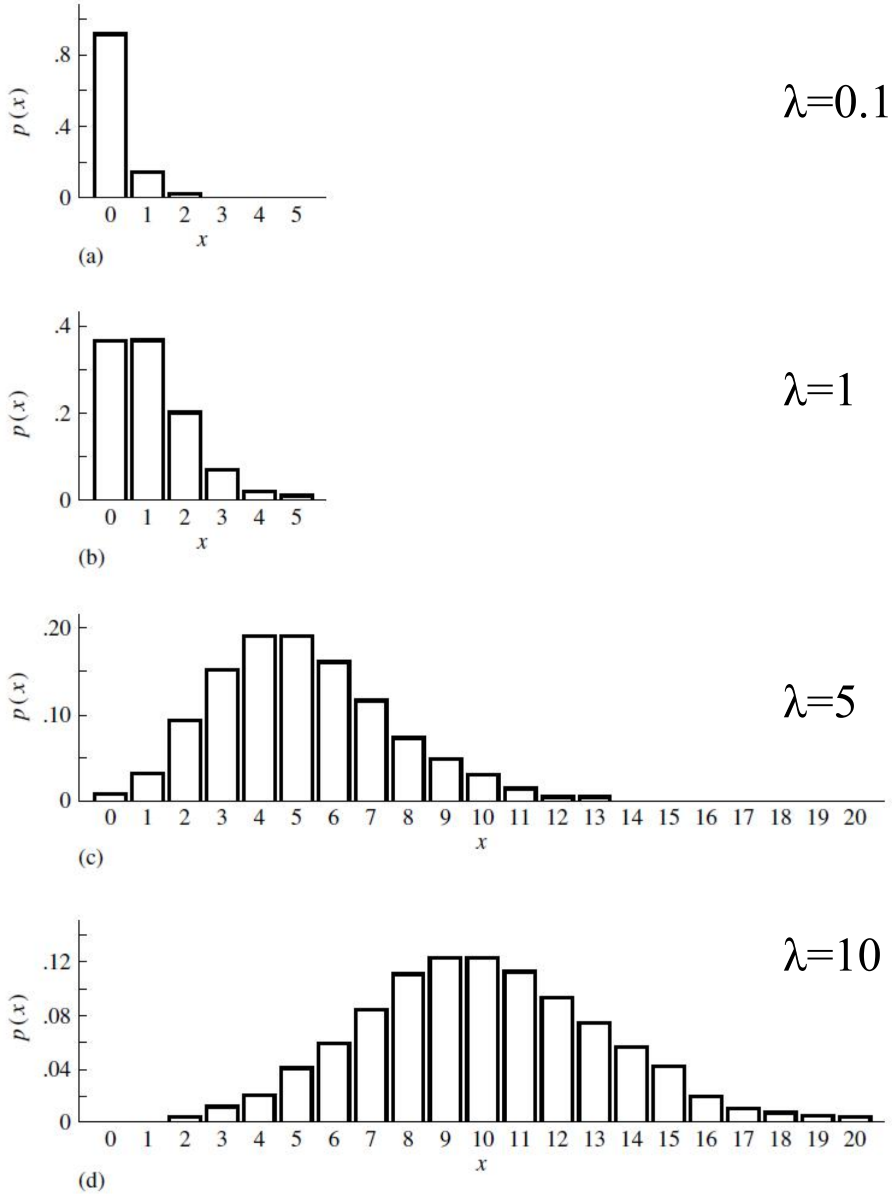
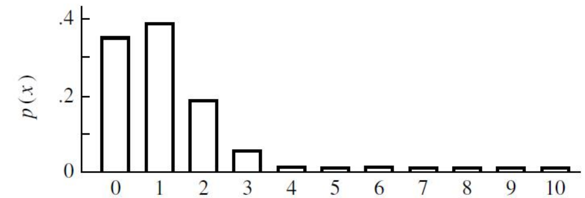


FIGURE **2.6** Poisson frequency functions, (a) $\lambda = .1$, (b) $\lambda = 1$, (c) $\lambda = 5$, (d) $\lambda = 10$.

As $n \to \infty$, $p \to 0$
Binomial distribution
→ Poisson distribution



Binomial: n=10, p=0.1

As $\lambda = np \to \infty$,
Poisson distribution
→ Gaussian / Normal distribution

A Glimpse at Poisson Process

Poisson distribution often arises from Poisson process for the distribution of random events in a set S (1-d:time, 2-d: a plane, 3-d: a volume of space).

互不相交子集

If $S_1, S_2, \ldots, S_n$ are disjoint subsets of S, then the numbers of events in these subsets, $N_1, N_2, \ldots, N_n$, are independent random variables that follow Poisson distributions with parameters $\lambda|S_1|, \lambda|S_2|, \ldots, \lambda|S_n|$, where $|S_i|$ is the measure of S_i (e.g. length, area, volume).

Crucial assumptions:

1. Events in disjoint subset are independent of each other;
2. Poisson parameter for a subset is proportional to the subset's size.

More in *stochastic processes.*

Random variables: continuous

Frequency / probability mass function → probability density function (**pdf**)

$$P(a < X < b) = \int_a^b f(x)\,dx \qquad \int_{-\infty}^{\infty} f(x)\, dx = 1$$

- *What is $P(X=c)$?*

Differential form:

$$P(x \le X \le x + dx) = f(x)\,dx.$$

Discrete cdf:

Continuous cdf:

$$F(x) = P(X \le x) \qquad \rightarrow \qquad F(x) = \int_{-\infty}^{x} f(u)\,du$$

Probability that X falls in an interval is:

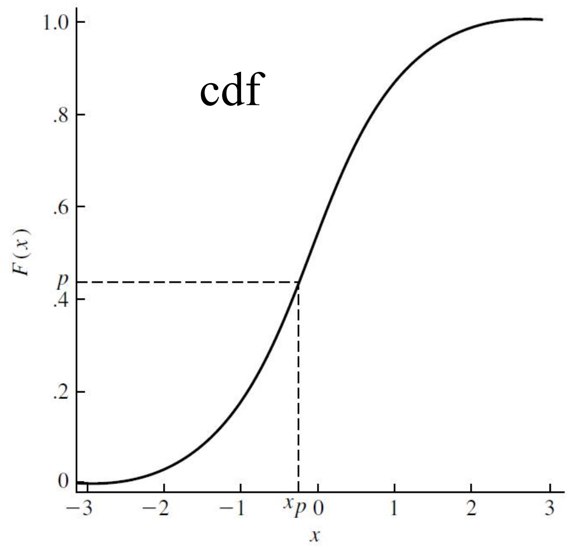
$$P(a \le X \le b) = \int_a^b f(x)\,dx = F(b) - F(a)$$

Quantile (分位数) and its application

The pth quantile x_p satisfies

$$F(x_p) = p, \text{ or } P(X \le x_p) = p.$$

50%-th quantile = **median**.

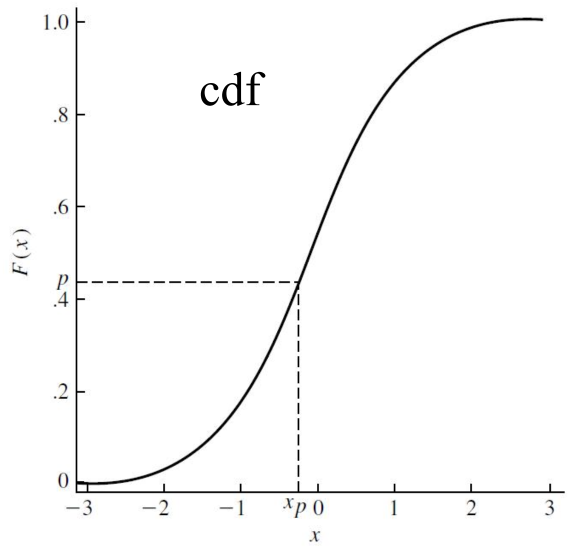


Quantile (分位数) and its application

The pth quantile x_p satisfies

$$F(x_p) = p, \text{ or } P(X \le x_p) = p.$$

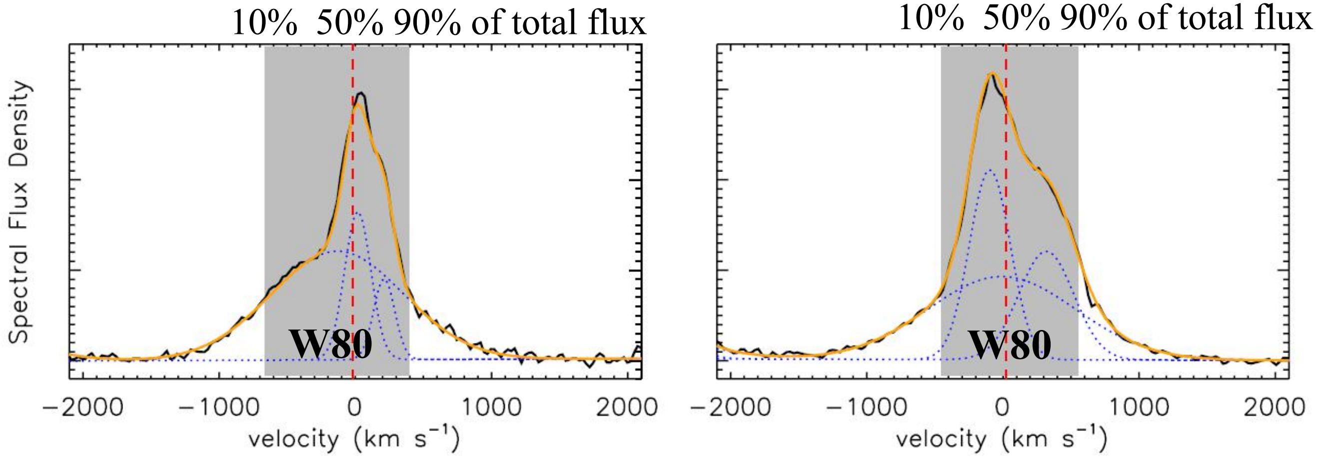
50%-th quantile = **median**.



Example from my research: [O III] 5007Å line

Why use quantiles?

- Quasar outflow physics hidden in the broad wings
- Need non-parametric measurement, because emission line has freaky profiles
- Quantile is sensitive to the flux under the broad wings



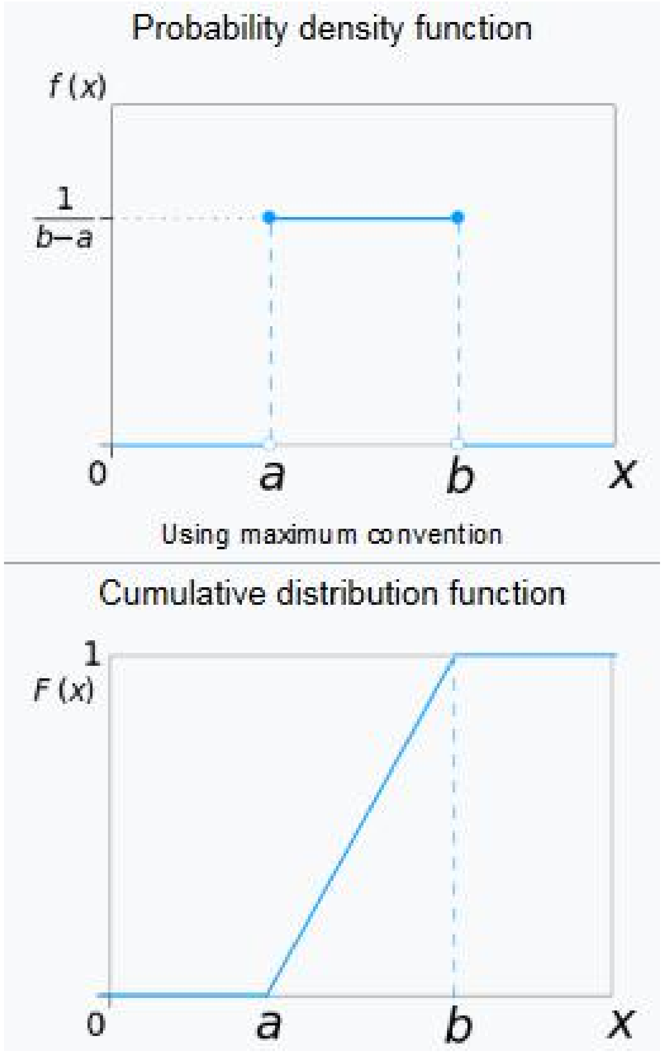
Liu et al. (2013)

Random variables: continuous

Uniform density:

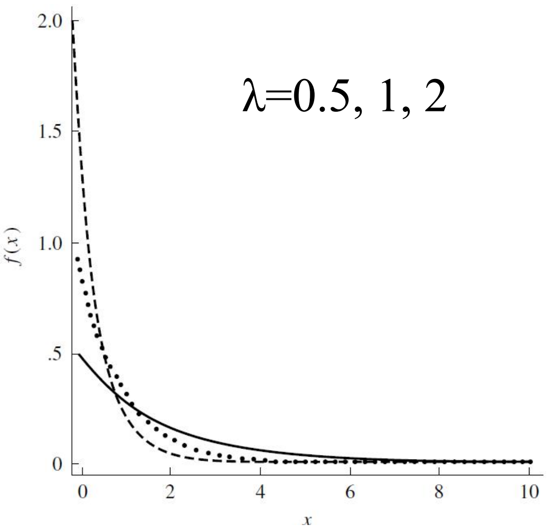
$$f(x) = \begin{cases} 1/(b-a), & a \le x \le b \\ 0, & x < a \text{ or } x > b \end{cases}$$

$$F(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 \le x \le 1 \\ 1, & x \ge 1 \end{cases}$$



Exponential distribution

pdf: $$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$



λ=0.5, 1, 2

cdf: $$F(x) = \int_{-\infty}^{x} f(u)\,du = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Often used to model lifetimes or waiting times.

Why?
Assume: at any time, the probability that an electronic component breaks down is a constant λ. ("***memoryless***", no aging effects)

Then: $$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Prove this in your homework.
(参见陈希孺《概率论与数理统计》2.1.3节，2009年版52页)

Exponential distribution

Reverse question:

1) An electronic component's lifetime is an exponential random variable,
2) it has lasted time *s*,
what is the probability that it will last at least more time *t*?

$$
\begin{aligned}
P(T > t + s \mid T > s) &= \frac{P(T > t + s \text{ and } T > s)}{P(T > s)} \\
&= \frac{P(T > t + s)}{P(T > s)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} \\
&= e^{-\lambda t}
\end{aligned}
$$

Result is independent of *s* – ***memoryless***, no aging effects
(therefore inapplicable on human lifetimes).

Exponential distribution

Memoryless character follows directly from Poisson process.

An event happens at t_0, events occur in time as a Poisson process (with λ for unit time). T=length of time until next event.

$$P(T > t) = P(\text{no events in } (t_0, t_0 + t))$$

During time t, Poisson distribution with parameter λt,

Don't be confused with the notations here [[$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, 2, \ldots$]]

No events, k=0, $e^{-\lambda t}$, an exponential distribution with λ.

The distribution of time until the 3rd event is similar, and is independent of the length of time between events 1 and 2.

The lengths of time between events of a Poisson process are independent, identically distributed, exponential random variables.

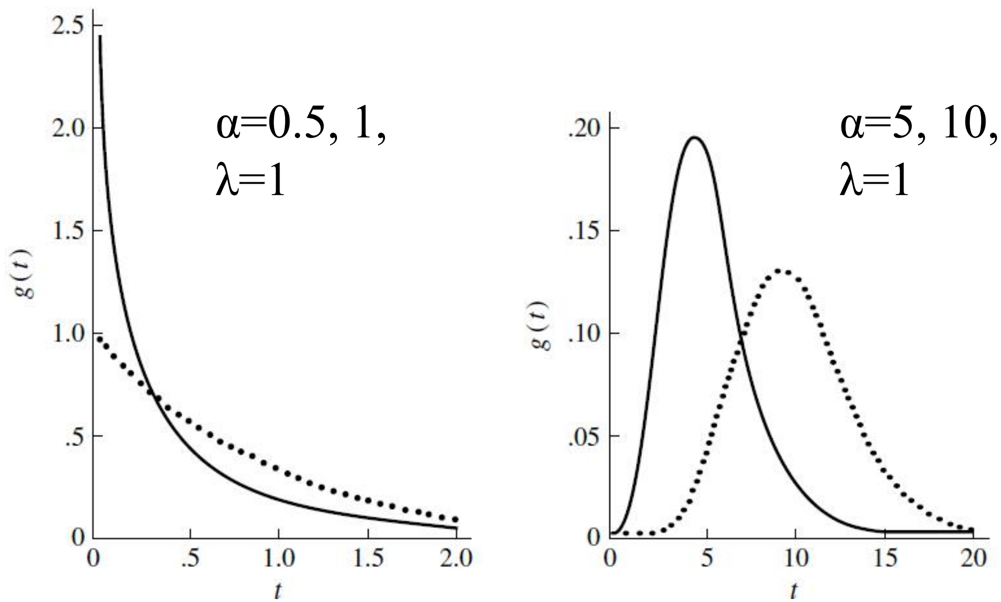
The Gamma density

Two parameters: α, λ, a flexible class for modeling nonnegative random variables,

$$g(t) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}, \qquad t \geq 0$$

Gamma function,

$$\Gamma(x) = \int_0^{\infty} u^{x-1} e^{-u}\, du, \qquad x > 0$$



Shape parameter
Scale parameter

The Gamma density

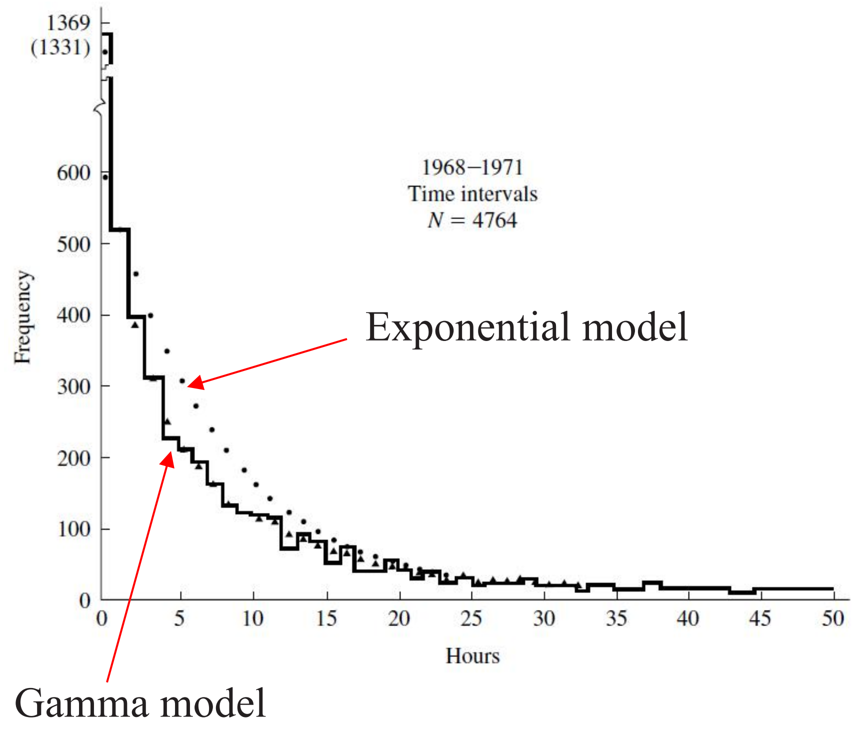
Example: (Udias & Rice, 1975)
The observed times separating a sequence of small earthquakes are fitted to a gamma density, and an exponential density.

Why does gamma density works better?

Because earthquakes are not memoryless!

Exponential model: knowing that an earthquake had not occurred in the last *t* time tells us nothing about the next *s* time.

Gamma model: a large likelihood that the next earthquake will immediately follow any given one, and this likelihood decreases monotonically with time.



Gaussian/Normal distribution

Why is normal distribution everywhere?

Because of the **central** limit theorem. (details later)

Roughly, if a random variable is the sum/average of a large number of independent random variable, it is approximately normally distributed.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty$$

$$X \sim N(\mu, \sigma^2)$$

Standard normal density has

$$\mu = 0 \text{ and } \sigma = 1$$

cdf has no close form.

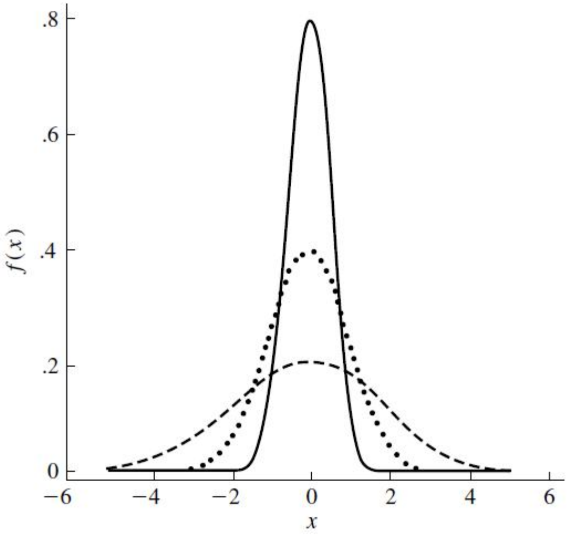
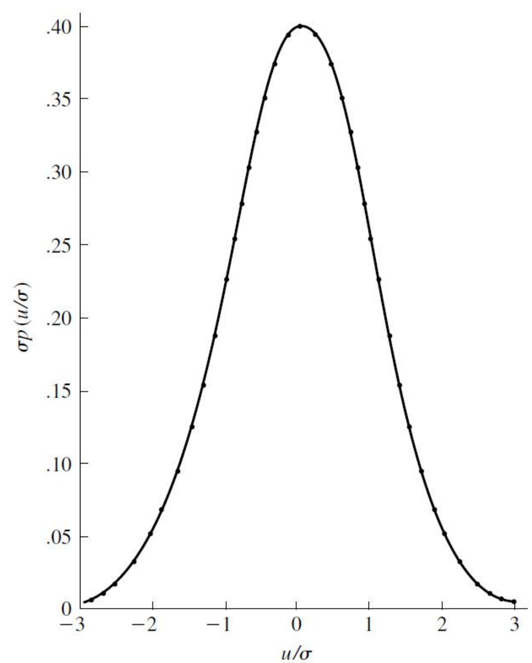


FIGURE 2.13 Normal densities, $\mu = 0$ and $\sigma = .5$ (solid), $\mu = 0$ and $\sigma = 1$ (dotted), and $\mu = 0$ and $\sigma = 2$ (dashed).

Gaussian/Normal distribution

Example:

Turbulent air flow is sometimes modeled as a random process. Since the velocity of the flow at any point is subject to the influence of a large number of random eddies in the neighborhood of that point, one might expect from the central limit theorem that the velocity would be normally distributed. Van Atta and Chen (1968) analyzed data gathered in a wind tunnel. Figure 2.16, taken from their paper, shows a normal distribution fit to 409,600 observations of one component of the velocity; the fit is remarkably good. ■

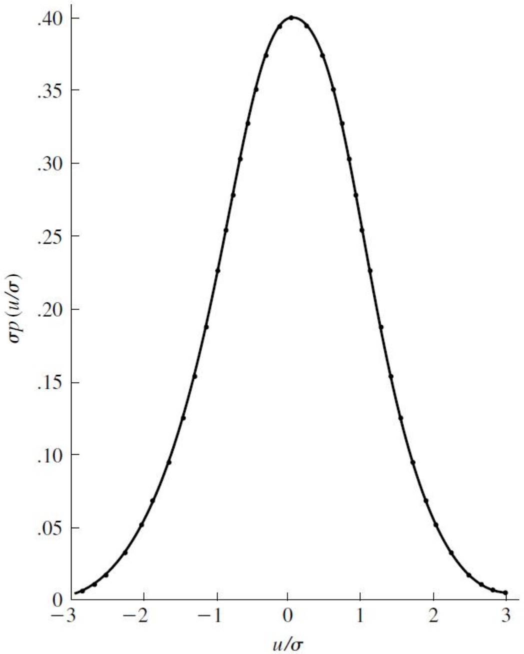


Gaussian/Normal distribution

Example:

Turbulent air flow is sometimes modeled as a random process. Since the velocity of the flow at any point is subject to the influence of a large number of random eddies in the neighborhood of that point, one might expect from the central limit theorem that the velocity would be normally distributed. Van Atta and Chen (1968) analyzed data gathered in a wind tunnel. Figure 2.16, taken from their paper, shows a normal distribution fit to 409,600 observations of one component of the velocity; the fit is remarkably good. ■

Q: why do you almost always find Gaussian noise on your CCD image?



Functions of a Random Variable

Q: We know the pdf of particle velocity, what is the pdf of the kinetic energy?

$f_X, F_X \rightarrow f_Y, F_Y$. Start with a simpler case.
Suppose that $X \sim N(\mu, \sigma^2)$ and $Y=aX+b$, $a>0$ ($a<0$ similar).

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(aX + b \le y) \\ &= P\left(X \le \frac{y-b}{a}\right) \\ &= F_X\left(\frac{y-b}{a}\right) \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_X\left(\frac{y-b}{a}\right) \\ &= \frac{1}{a} f_X\left(\frac{y-b}{a}\right) \end{aligned}$$

after substitution,

$$f_Y(y) = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-b-a\mu}{a\sigma}\right)^2\right]$$

PROPOSITION A

If $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$, then $Y \sim N(a\mu + b, a^2\sigma^2)$.

Functions of a Random Variable

Q: We know the pdf of particle velocity, what is the pdf of the kinetic energy?

Find the density of $X=Z^2$, where $Z\sim N(0,1)$.

$$\begin{aligned} F_X(x) &= P(X \le x) \\ &= P(-\sqrt{x} \le Z \le \sqrt{x}) \\ &= \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) \end{aligned}$$

standard normal distribution
cdf to pdf $\Phi'(x) = \phi(x)$

Differentiating (we circumvent cdf because it has no closed form),

$$\begin{aligned} f_X(x) &= \tfrac{1}{2}x^{-1/2}\phi(\sqrt{x}) + \tfrac{1}{2}x^{-1/2}\phi(-\sqrt{x}) \\ &= x^{-1/2}\phi(\sqrt{x}) \end{aligned}$$

Finally,

$$f_X(x) = \frac{x^{-1/2}}{\sqrt{2\pi}}e^{-x/2}, \qquad x \ge 0$$

$$g(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)}t^{\alpha-1}e^{-\lambda t}, \qquad t \ge 0$$

$$\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}.$$

Gamma density, with $\alpha=\lambda=1/2$.
Chi-square density with 1 degree of freedom.

Functions of a Random Variable

We always go through the same steps:
Find the cdf of Y, diffcrentiate it to find pdf, specify in what region it holds.

PROPOSITION B

Let X be a continuous random variable with density $f(x)$ and let $Y = g(X)$ where g is a differentiable, strictly monotonic function on some interval I. Suppose that $f(x) = 0$ if x is not in I. Then Y has the density function

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

for y such that $y = g(x)$ for some x, and $f_Y(y) = 0$ if $y \neq g(x)$ for any x in I. Here g^{-1} is the inverse function of g; that is, $g^{-1}(y) = x$ if $y = g(x)$. ■

However, for any specific problem, proceeding from scratch is usually easier.

Generating pseudorandom numbers

PROPOSITION **C**

Let $Z = F(X)$; then Z has a uniform distribution on [0, 1].

Proof

$$P(Z \leq z) = P(F(X) \leq z) = P(X \leq F^{-1}(z)) = F(F^{-1}(z)) = z$$

This is the uniform cdf. ■

Generating pseudorandom numbers

PROPOSITION **C**

Let $Z = F(X)$; then Z has a uniform distribution on [0, 1].

Proof

$$P(Z \le z) = P(F(X) \le z) = P(X \le F^{-1}(z)) = F(F^{-1}(z)) = z$$

This is the uniform cdf. ■

PROPOSITION **D**

Let U be uniform on [0, 1], and let $X = F^{-1}(U)$. Then the cdf of X is F.

Proof

$$P(X \le x) = P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x)$$ ■

To generate random variables with cdf F, simply apply F^{-1} to uniform random numbers.

Generating pseudorandom numbers

Example.
Generating random variables from an exponential distribution.

When simulating large queueing networks to assess the performance, one needs to generate random time intervals between customer arrivals, which might be exponentially distributed. (***Why?***)

cdf: $$F(t) = 1 - e^{-\lambda t}$$

F^{-1}:
$$e^{-\lambda t} = 1 - x$$
$$-\lambda t = \log(1 - x)$$
$$t = -\log(1 - x)/\lambda$$

So if U is uniform on [0,1], then $T = -\log(1 - U)/\lambda$ is the result.

Actually $V = 1 - U$ is also uniformly distributed on [0, 1], so finally

$$T = -\log(V)/\lambda, \text{ where } V \text{ is uniform on } [0, 1].$$

Review: Conditional probability & Bayes' rule

DEFINITION

Let A and B be two events with $P(B) \neq 0$. The conditional probability of A given B is defined to be

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad \blacksquare$$

LAW OF TOTAL PROBABILITY

Let $B_1, B_2, \ldots, B_n$ be such that $\bigcup_{i=1}^{n} B_i = \Omega$ and $B_i \cap B_j = \emptyset$ for $i \neq j$, with $P(B_i) > 0$ for all i. Then, for any event A,

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i)$$

BAYES' RULE

Let A and $B_1, \ldots, B_n$ be events where the B_i are disjoint, $\bigcup_{i=1}^{n} B_i = \Omega$, and $P(B_i) > 0$ for all i. Then

$$P(B_j \mid A) = \frac{P(A \mid B_j) P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i) P(B_i)}$$

The proof of Bayes' rule follows exactly as in the preceding discussion. $\blacksquare$

Laplace's law of succession

Suppose that the sun has risen *n* times in succession; what is the probability that it will rise once more?

Pierre-Simon Laplace (1749–1827)

Laplace's law of succession

Suppose that the sun has risen n times in succession; what is the probability that it will rise once more?

Assume the *a priori* probability for a sunrise on any day is a constant. Due to our total ignorance it will be assumed to take all possible values in $[0, 1]$ with equal likelihood.

This probability is treated as a random variable ξ uniformly distributed over $[0, 1]$. Thus ξ has the density function f such that $f(p) = 1$ for $0 \le p \le 1$.

$$P(p \le \xi \le p + dp) = dp, \quad 0 \le p \le 1.$$

Assume sunrises are independent events, if the true value of ξ is p, then the probability of n successive sunrises is p^n.

$$P(S^n \mid \xi = p) = p^n.$$

Laplace's law of succession

Suppose that the sun has risen *n* times in succession; what is the probability that it will rise once more?

Law of total probability:

$$P(S^n) = \sum_{0 \le p \le 1} P(\xi = p) P(S^n \mid \xi = p).$$

Pass from the sum into an integral,

$$P(S^n) = \int_0^1 P(S^n \mid \xi = p)\, dp = \int_0^1 p^n\, dp = \frac{1}{n+1}.$$

Apply it for both *n* and *n*+1, then take the ratio,

$$P(S^{n+1} \mid S^n) = \frac{P(S^n S^{n+1})}{P(S^n)} = \frac{P(S^{n+1})}{P(S^n)} = \frac{\frac{1}{n+2}}{\frac{1}{n+1}} = \frac{n+1}{n+2}.$$

Problem set #1

1. Do you agree with Laplace's result? No matter your answer is yes or no, elaborate your comments (e.g. meaning, assumptions, derivation, interpretation, Bayesian spirit, …, if any). Use your critical thinking!
2. Find the distribution of $Y=X-r$, where X is a negative binomial random variable. Then write down the Taylor expansion of $(1-x)^{-r}$. Now, do you understand why this distribution is called "negative binomial"?
3. Verify that Bayes' rule indeed renders 7.5% in the cancer research mentioned in Lecture 1 (see slide pp. 27).
4. Prove that under the memoryless assumption, the lifetime of electronic components follow the exponential distribution.
5. Create an exponential distribution using the F^{-1} technique. Plot a histogram of the data you generate overlaid with the expected pdf to show you are successful. Do the same thing on another 1-2 of any continuous distributions.