Random Processes and Normal Distributions

SU 5050
LECTURE 6

JESSICA L. MCCARTY, PH.D.

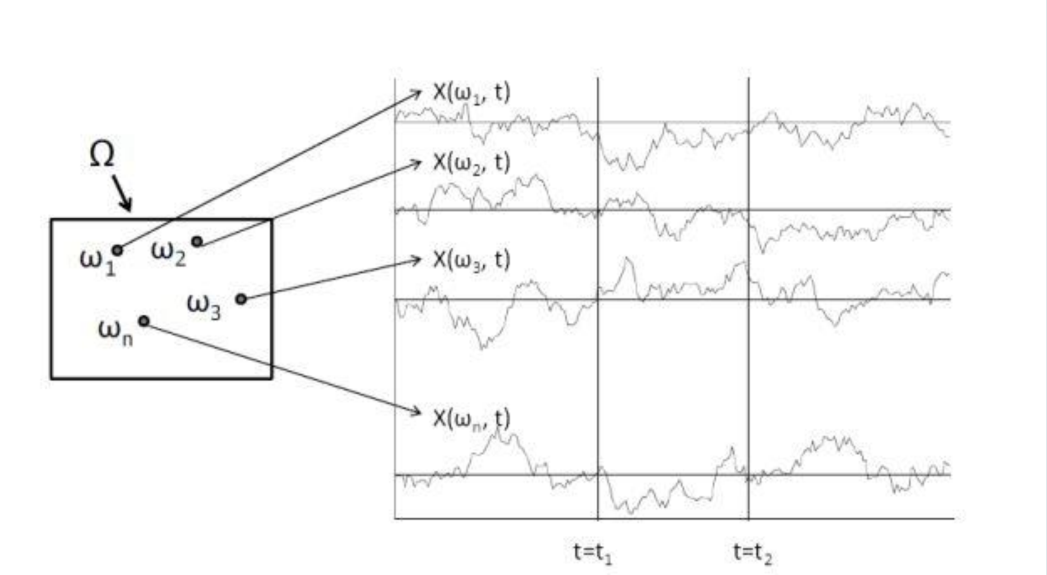
Random Sample

- A random sample is a sample in which each individual or object in the population has an equal chance of being selected.

- A *random process* – a collection of random variables – is often called a *stochastic process* in probability theory.

Random processes

- A probability system, which is composed of a sample space, a set of real-valued time-indexed functions, and a probability measure, is called a ***random process*** or a ***stochastic process***.



Random Processes

- The individual time functions of the random process X(t) are called **sample functions**.

- By definition, a random process implies the existence of an infinite number of random variables, one for each t in some range.

Classification of Random Processes

Discrete-Time vs. Continuous-Time Processes

- A discrete-time random process or a random sequence, denoted as X_t or X_k.
- A continuous-time random process, denoted as $X(t)$
- The term **time series** is used synonymously with discrete-time random process

Classification of Random Processes

- **State space** = a set of possible values that a random process, discrete or continuous in time, may take on.

Examples of **discrete states** :

A Bernoulli trial: S={ s, f } or S={ 0, 1}

A simple random walk: $S = \{0, \pm 1, \pm 2, \pm 3, \ldots \}$ (step size)

The price of a stock: $S = \{0, 1, 2, \ldots \}$ (unit price)

Examples of **continuous states**:

The temperature as a function of time: $S = (-\infty, \infty)$

Gaussian process: $S = (-\infty, \infty)$

Brownian motion or Wiener process, a limit of the random walk : $S = (-\infty, \infty)$

Inter-arrival time of a Poisson process: $S = [0, \infty)$

- Digital technology transforms continuous-time/continuous-space process to a discrete-time/discrete-space process (DVDs, mp3s, podcasts).

Independent vs. Dependent Processes

Suppose we arbitrarily choose n time instants and consider the joint distribution function $F_{\boldsymbol{X}}(\boldsymbol{x}, \boldsymbol{t})$ of the set of random variables $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ where $X_i = X(t_i); i = 1, 2, \ldots, n$. If this distribution function factors into the product:

$$F_{\mathbf{X}}(\boldsymbol{x}; \boldsymbol{t}) \triangleq F_{X_1 X_2 \cdots X_n}(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n)$$
$$= F_{X_1}(x_1; t_1) F_{X_2}(x_2; t_2) \cdots F_{X_n}(x_n; t_n),$$

for any finite n and for any choice of the instants $\boldsymbol{t}$, we say $X(t)$ is an **independent process**.

Examples of **independent** processes:

A Bernoulli trials

Step sizes of a random walk (i.e., the difference sequence of a random walk)

White noise (i.e., the power spectral is flat for all frequencies)

Interarrival times of a Poisson process

Many examples of random sequences X_k or X_n discussed in Chapter 11.

Examples of **dependent** process

A **random walk**

Brownian motion (integration of white noise)

A **Makov** process (discrete-time, continuous-time)

Packet traffic over LAN is known to have long-range dependency (LRD)

Discrete-Time Markov Chain (DTMC)

A discrete-time random process $\{X_k\}$ is called a **simple Markov chain**, if X_{k+1} is Independent of $X_1, X_2, \ldots, X_{k-1}$ in case X_k is known

i.e., if X_{k+1} depends on its past only through its most recent value X_k.

A Markov chain of **order *h*** is a sequence in which X_k depends on its past only through its ***h* previous values**, X_{k-1}, X_{k-2}, … , X_{k-h}.

$$p(x_k | x_{k-i};\ i \geq 1) = p(x_k | x_{k-1}, x_{k-2}, \ldots, x_{k-h}).$$

A Markov chain of order *h* defined over state space S can be transformed into a simple Markov chain by defining the state space $S^h = S \times S \cdots \times S$, the *h*-times Cartesian product of *S* with itself.

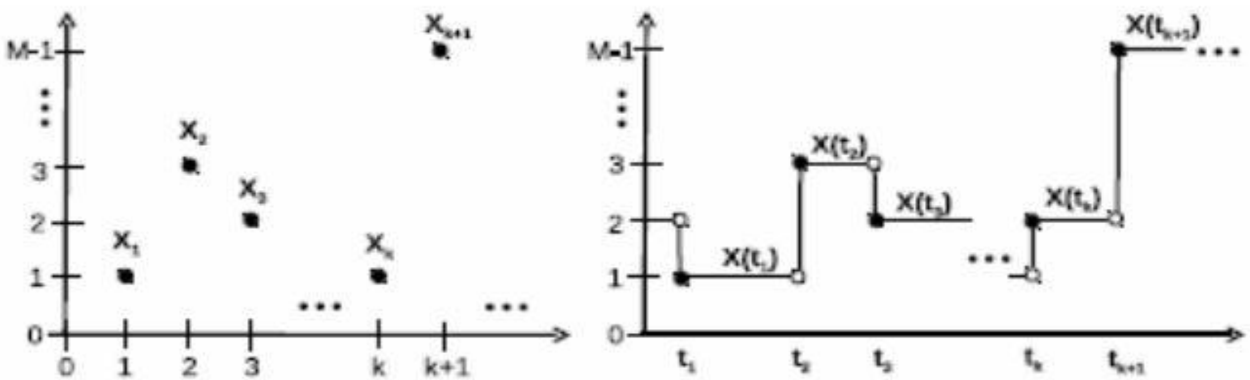
Discrete-Time Markov Chain (DTMC)

Markov chain models and related **hidden Markov models** (**HMM**s) are used in a variety of fields, including linguistic models for speech recognition, DNA and protein sequences, network traffic, etc. (cf. Chapters 1 and 20).

The simple Markov chain defined above is often referred to as a **discrete-time Markov chain (DTMC)**.

If there are M different states, we can label them, without loss of generality, by integers 0, 1, 2, M-1, i.e.,

$$\mathcal{S} = \{0, 1, 2, \ldots, M-1\},$$



(a) DTMC $\{X_k\}$ sample path. (b) CTMC $X(t)$ sample path.

Continuous-Time Markov Chain (CTMC)

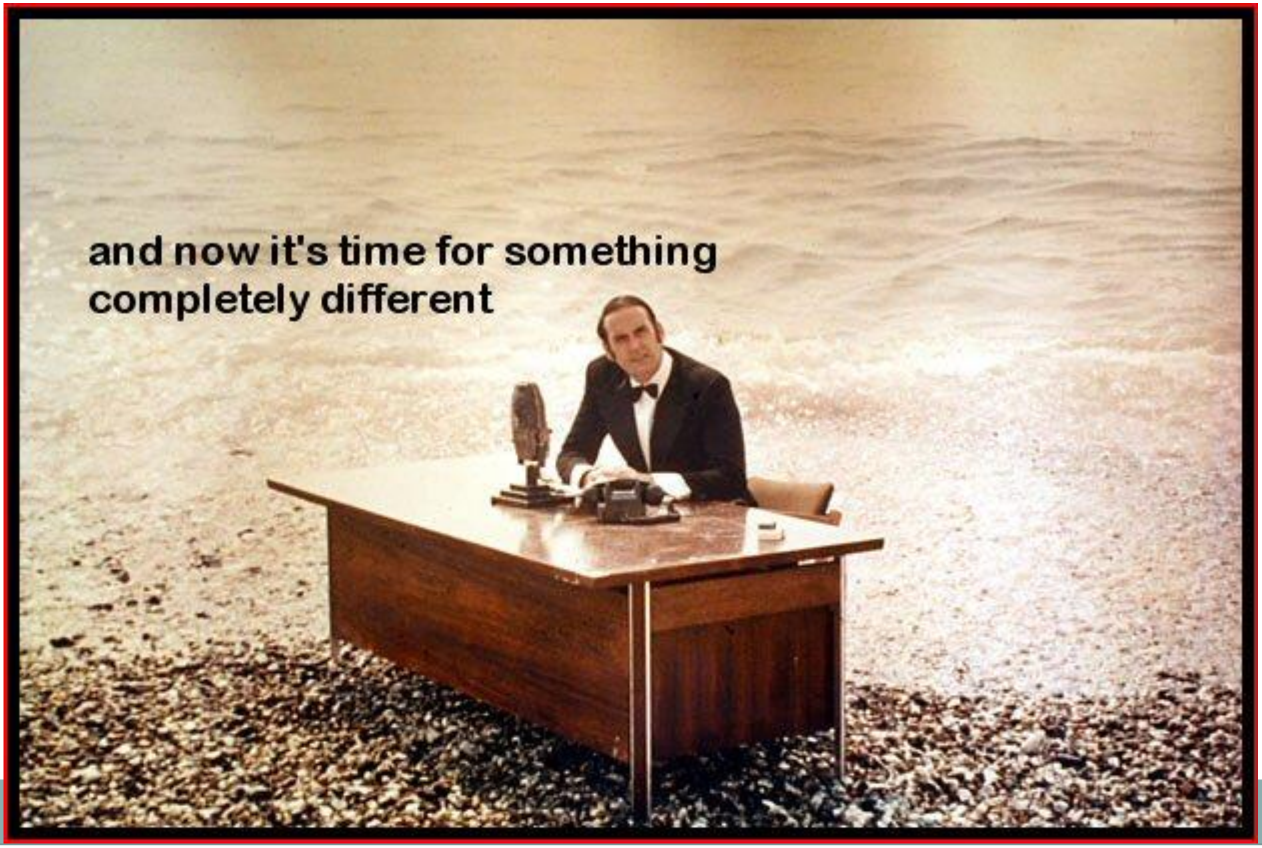
For a given DTMC $\{X_k\}$ we can construct a **continuous-time Markov chain (CTMC)** $X(t)$

$$X(t) = i, \quad \text{for } t_k \le t < t_{k+1}, \quad \text{where } i = X(t_k), \quad \text{and } X(t_{k+1}) = j (\neq i).$$

and let the interval $\tau_k \triangleq t_{k+1} - t_k$ be **exponentially distributed** with mean λ_k^{-1}.

The future behavior of $X(t); t \ge t_n$ depends on its past $X(s); -\infty < s < t_n$ only through its current state $X(t_n) = i \in S$, because of the **memoryless property** of the exponential distribution.

For a **Poisson arrival** process, let $X(t)$ be the cumulative number of arrivals (or births) up to time t. This **counting process** is a CTMC.

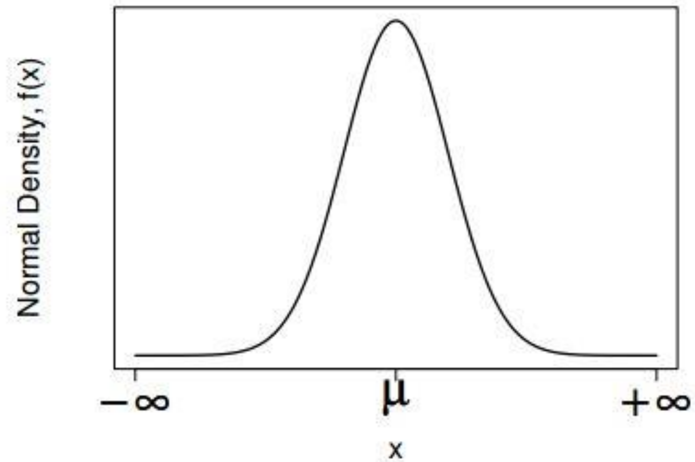

and now it's time for something completely different

Let's Remember Normal Distribution

- A probability distribution for continuous data
- Characterized by a symmetric bell-shaped curve (Gaussian curve)

  - Symmetric about its mean μ
- Under **certain conditions**, can be used to approximate Binomial(n,p) distribution
  - $np>5$
  - $n(1-p)>5$

Normal Distribution



- Takes on values between $-\infty$ and $+\infty$
- Mean = Median = Mode
- Area under curve equals 1

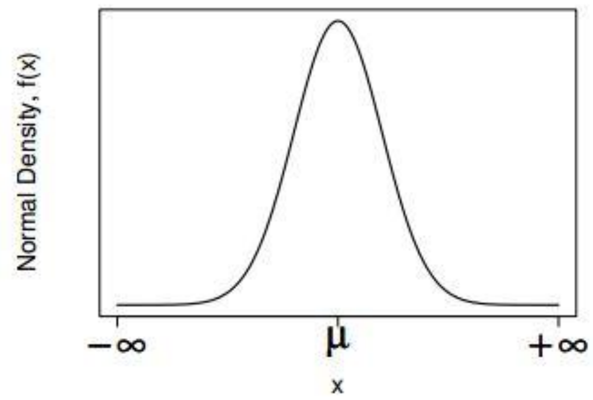
Notation for Normal random variable: $X \sim N(\mu, \sigma^2)$

- Parameters

  μ = mean

  σ = standard deviation

Normal Probability Density Function (PDF)



The normal probability density function for $X \sim N(\mu, \sigma^2)$ is:

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

Note: $\pi \approx 3.14$ and $e \approx 2.72$ are mathematical constants

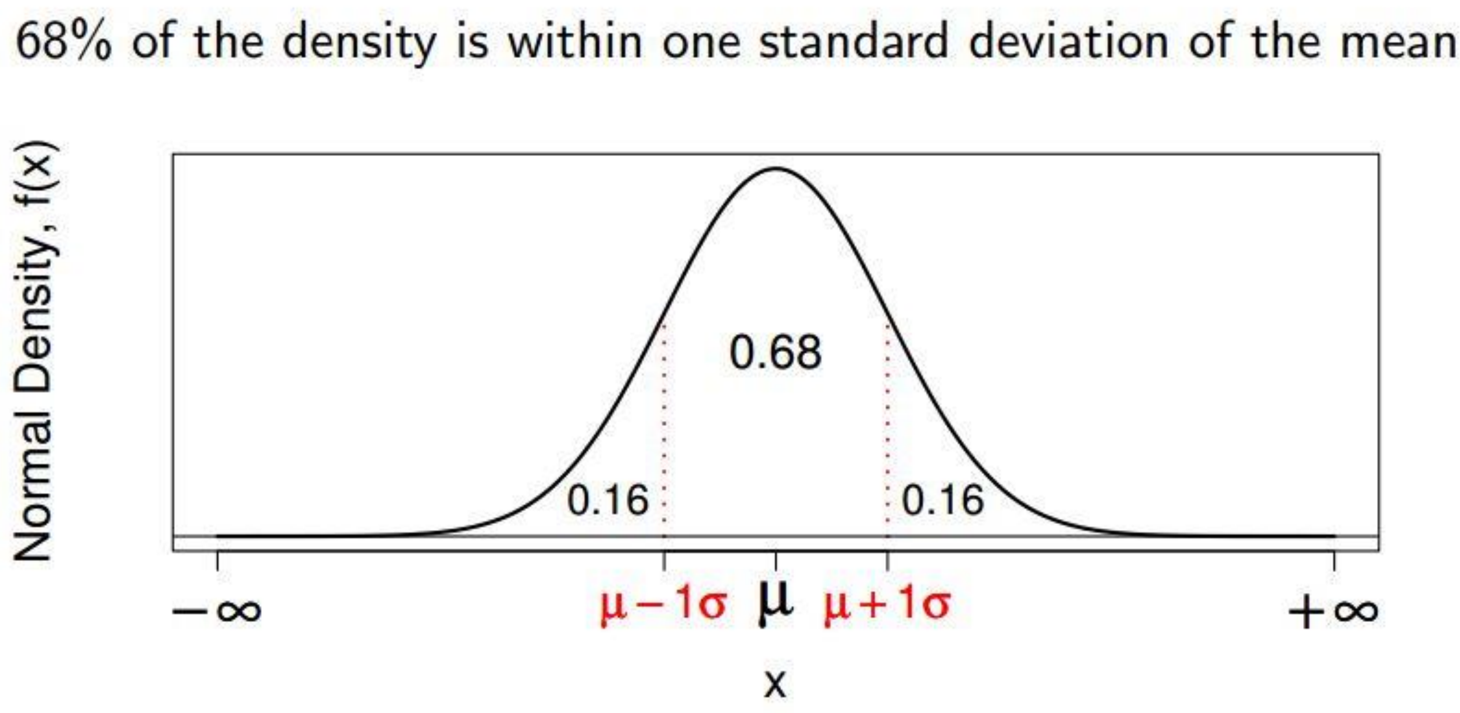
Standard Normal

- Definition: a Normal distribution $N(\mu, \sigma^2)$ with parameters $\mu = 0$ and $\sigma = 1$
- Its density function is written as:

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

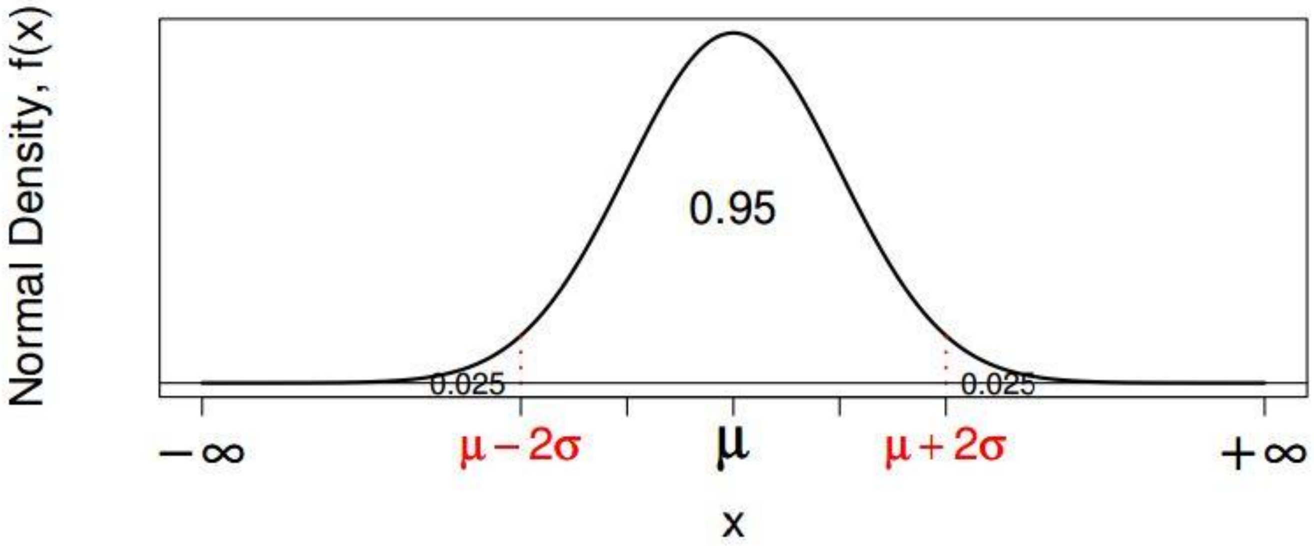
- We typically use the letter Z to denote a standard normal random variable ($Z \sim N(0, 1)$)
- *Important!* We use the standard normal all the time because if $X \sim N(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim N(0, 1)$
- This process is called "standardizing" a normal random variable

Normal Distribution Rules: Rule #1



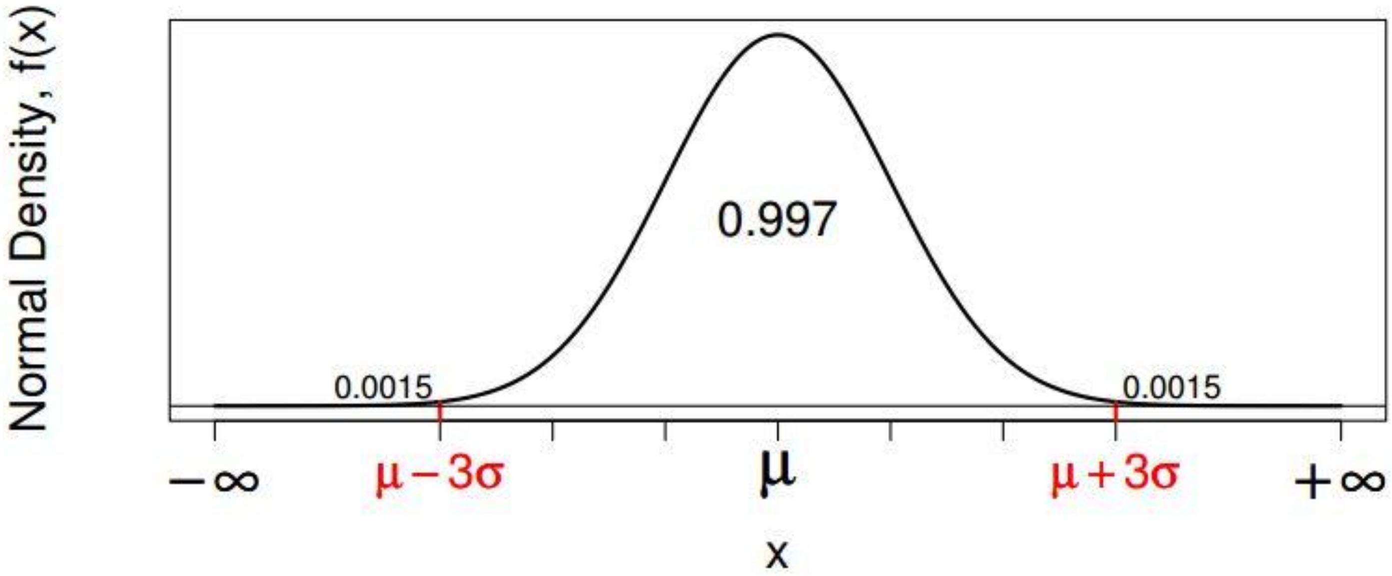
Rule #2

95% of the density is within two standard deviations of the mean

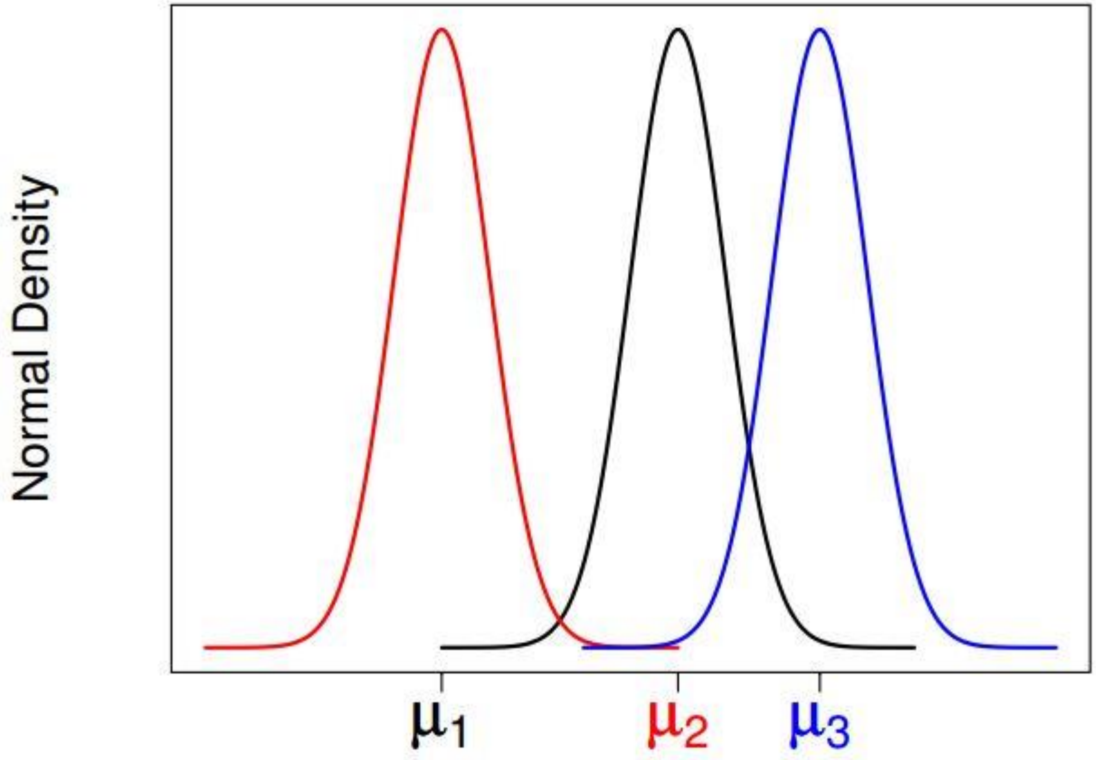


Rule #3

99.7% of the density is within three standard deviations of the mean



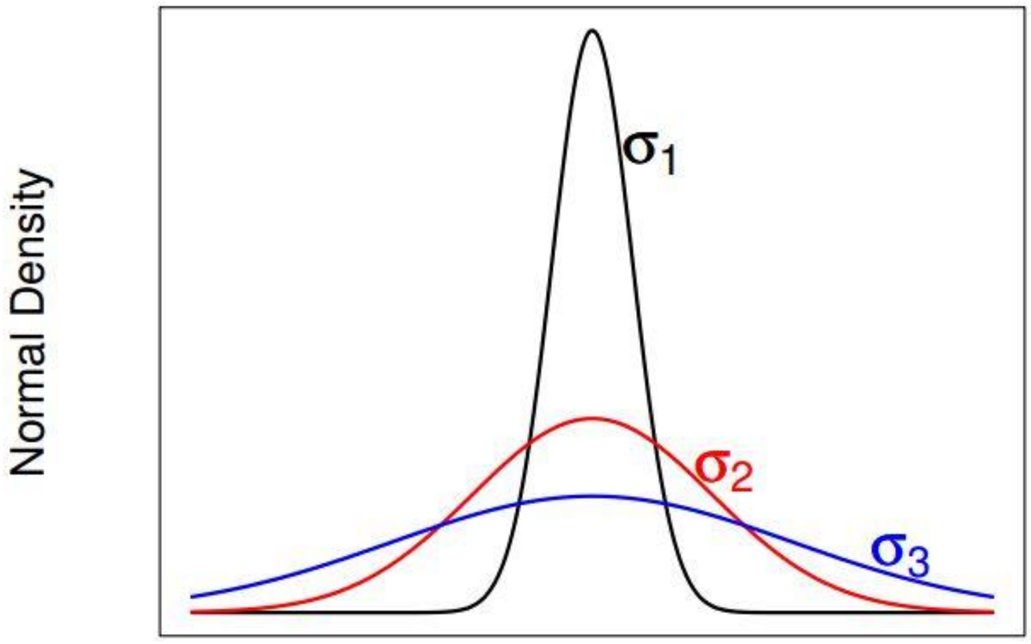
They Can Have Different Means…



Three normal distributions with different means

$\mu_1 < \mu_2 < \mu_3$

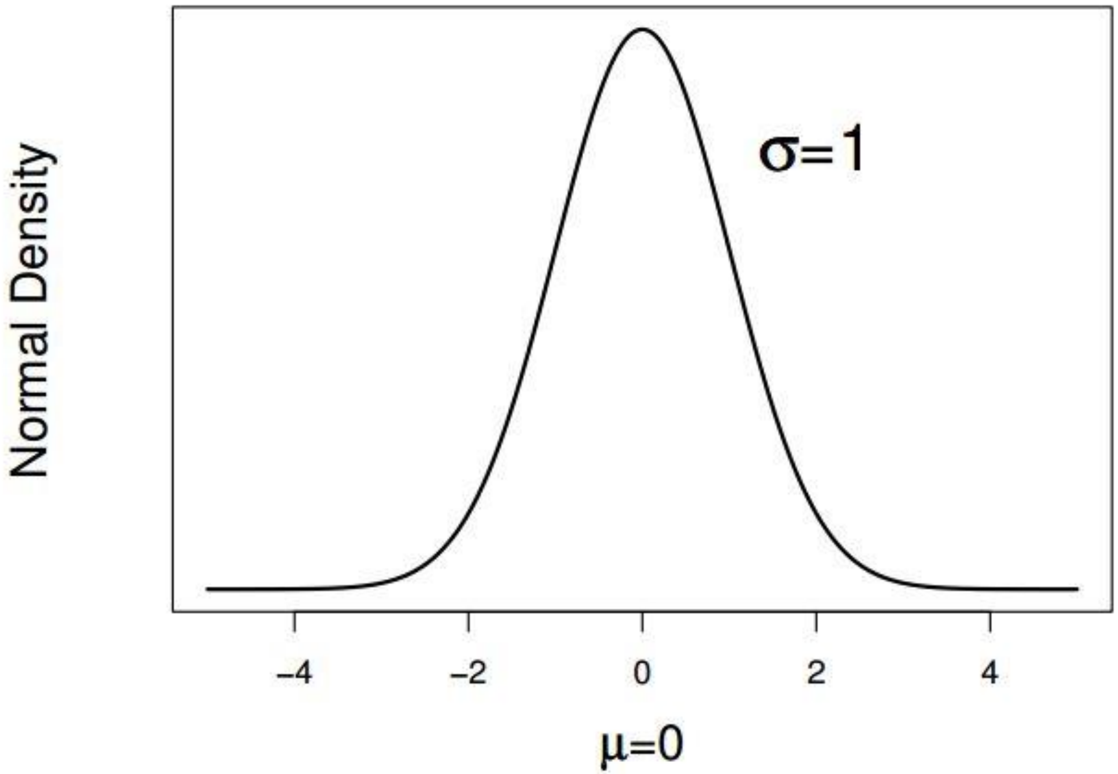
… and Different Standard Deviations



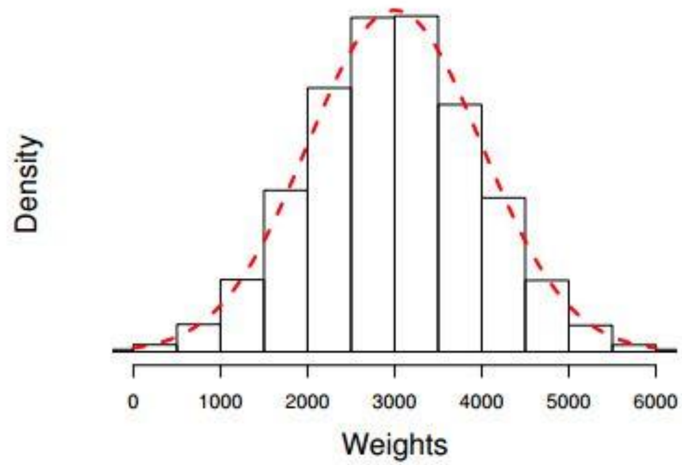
Three normal distributions with different standard deviations

$\sigma_1 < \sigma_2 < \sigma_3$

Ideal: Standard Normal N(0,1)



Example: Birth weights (grams) of infants

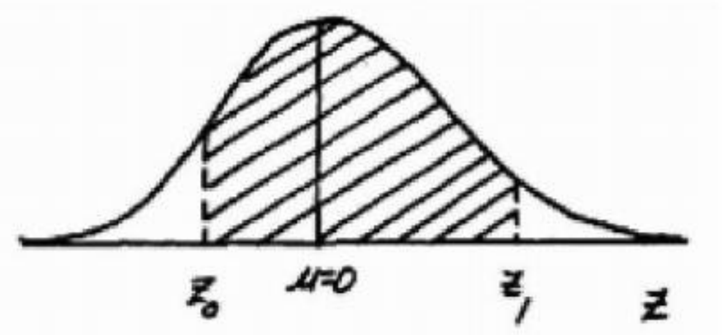


- Continuous data
- Mean $=$ Median $=$ Mode $= 3000 = \mu$
- Standard deviation $= 1000 = \sigma$
- The area under the curve represents the probability (proportion) of infants with birthweights between certain values

Calculate Normal Probabilities

We are often interested in the probability that z takes on values between z_0 and z_1

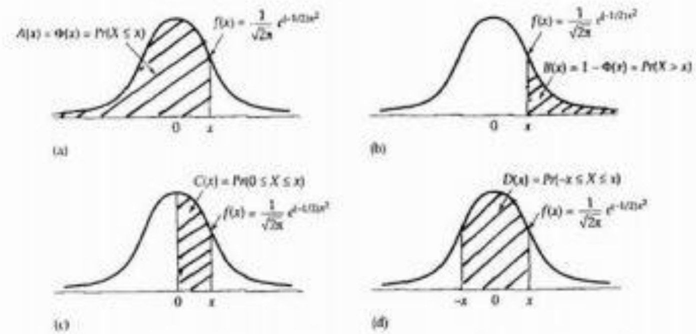
$$P(z_0 \leq z \leq z_1) = \int_{z_0}^{z_1} \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2/2} dz$$



How do we calculate this probability?

- Equivalent to finding area under the curve
- Continuous distribution, so we cannot use sums to find probabilities
- Performing the integration is not necessary since tables and computers are available

Z Tables



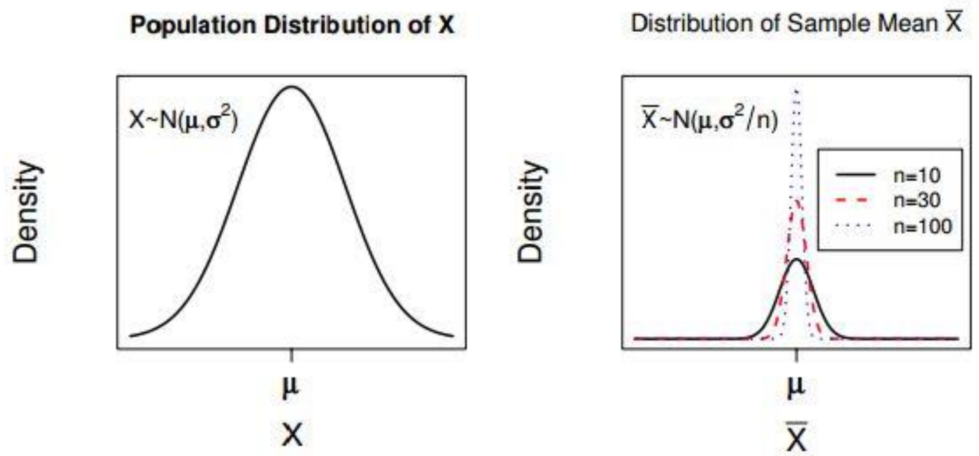
x	A^a	B^b	C^c	D^d	x	A	B	C	D
0.0	.5000	.5000	.0	.0	0.32	.6255	.3745	.1255	.2510
0.01	.5040	.4960	.0040	.0080	0.33	.6293	.3707	.1293	.2586
0.02	.5080	.4920	.0080	.0160	0.34	.6331	.3669	.1331	.2661
0.03	.5120	.4880	.0120	.0239	0.35	.6368	.3632	.1368	.2737
0.04	.5160	.4840	.0160	.0319	0.36	.6406	.3594	.1406	.2812
0.05	.5199	.4801	.0199	.0399	0.37	.6443	.3557	.1443	.2886
0.06	.5239	.4761	.0239	.0478	0.38	.6480	.3520	.1480	.2961
0.07	.5279	.4721	.0279	.0558	0.39	.6517	.3483	.1517	.3035
0.08	.5319	.4681	.0319	.0638	0.40	.6554	.3446	.1554	.3108
0.09	.5359	.4641	.0359	.0717	0.41	.6591	.3409	.1591	.3182
0.10	.5398	.4602	.0398	.0797	0.42	.6628	.3372	.1628	.3255
0.11	.5438	.4562	.0438	.0876	0.43	.6664	.3336	.1664	.3328
0.12	.5478	.4522	.0478	.0955	0.44	.6700	.3300	.1700	.3401
0.13	.5517	.4483	.0517	.1034	0.45	.6736	.3264	.1736	.3473
0.14	.5557	.4443	.0557	.1113	0.46	.6772	.3228	.1772	.3545
0.15	.5596	.4404	.0596	.1192	0.47	.6808	.3192	.1808	.3616
0.16	.5636	.4364	.0636	.1271	0.48	.6844	.3156	.1844	.3688
0.17	.5675	.4325	.0675	.1350	0.49	.6879	.3121	.1879	.3759
0.18	.5714	.4286	.0714	.1428	0.50	.6915	.3085	.1915	.3829
0.19	.5753	.4247	.0753	.1507	0.51	.6950	.3050	.1950	.3899
0.20	.5793	.4207	.0793	.1585	0.52	.6985	.3015	.1985	.3969
0.21	.5832	.4168	.0832	.1663	0.53	.7019	.2981	.2019	.4039
0.22	.5871	.4129	.0871	.1741	0.54	.7054	.2946	.2054	.4108
0.23	.5910	.4090	.0910	.1819	0.55	.7088	.2912	.2088	.4177
0.24	.5948	.4052	.0948	.1897	0.56	.7123	.2877	.2123	.4245
0.25	.5987	.4013	.0987	.1974	0.57	.7157	.2843	.2157	.4313
0.26	.6026	.3974	.1026	.2051	0.58	.7190	.2810	.2190	.4381
0.27	.6064	.3936	.1064	.2128	0.59	.7224	.2776	.2224	.4448
0.28	.6103	.3897	.1103	.2205	0.60	.7257	.2743	.2257	.4515
0.29	.6141	.3859	.1141	.2282	0.61	.7291	.2709	.2291	.4581
0.30	.6179	.3821	.1179	.2358	0.62	.7324	.2676	.2324	.4647
0.31	.6217	.3783	.1217	.2434	0.63	.7357	.2643	.2357	.4713

(continued on next page)

Python

- Statistics modules/libraries:
  - http://statsmodels.sourceforge.net/stable/
  - http://scikits.appspot.com/statsmodels

Sampling Distribution of $\bar{X}$



When sampling from a **normally** distributed population

- $\bar{X}$ will be normally distributed
- The mean of the distribution of $\bar{X}$ is equal to the true mean μ of the population from which the samples were drawn
- The variance of the distribution is σ^2/n, where σ^2 is the variance of the population and n is the sample size
- We can write: $\bar{X} \sim N(\mu, \sigma^2/n)$

When sampling from a population whose distribution is not **normal** and the sample size is **large**, use the Central Limit Theorem

The Central Limit Theorem (CLT)

Given a population of **any** distribution with mean, μ, and variance, σ^2, the sampling distribution of $\bar{X}$, computed from samples of size n from this population, will be **approximately** $N(\mu, \sigma^2/n)$ when the sample size is large

- In general, this applies when $n \geq 25$
- The approximation of normality becomes better as n increases

Normal distribution = Gaussian distribution

Carl Friedrich Gauss (1777-1855) on the German 10 Mark Note

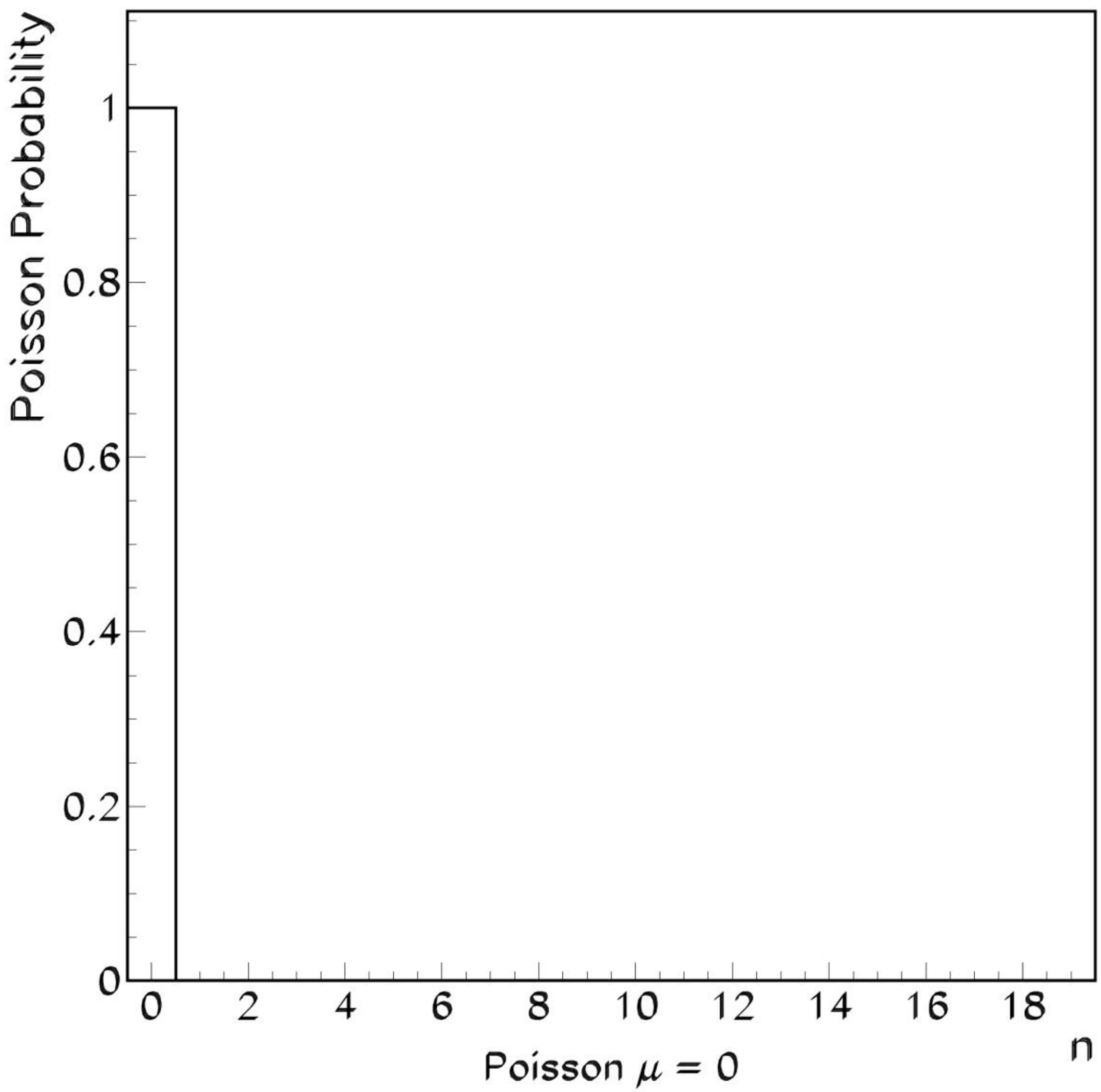
3 Important Probability Distributions

- Binomial : Result of experiment can be described as yes/no or success/failure outcome of a trial. *Probability of obtaining success is known*.
- Poisson: Predicts outcome of "counting experiments" where the expected number of counts is **small**. *Examples: Radiation with Geiger counter, bubbles in a bubble chamber track*.
- Gaussian: Predicts outcome of "counting experiments" where the expected number of counts is **large**.

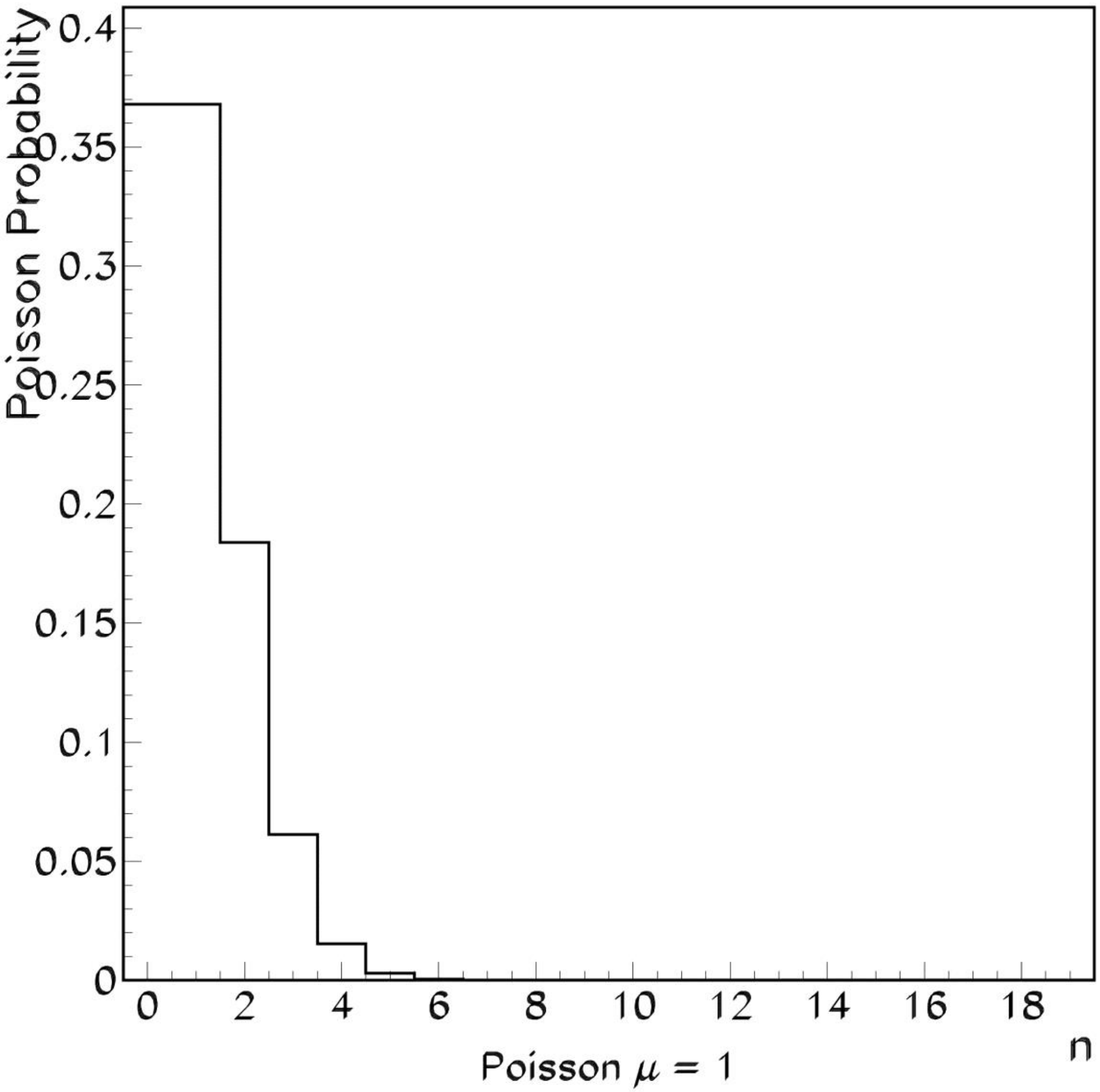
Poisson

Simon Denis Poisson
(1781-1840)

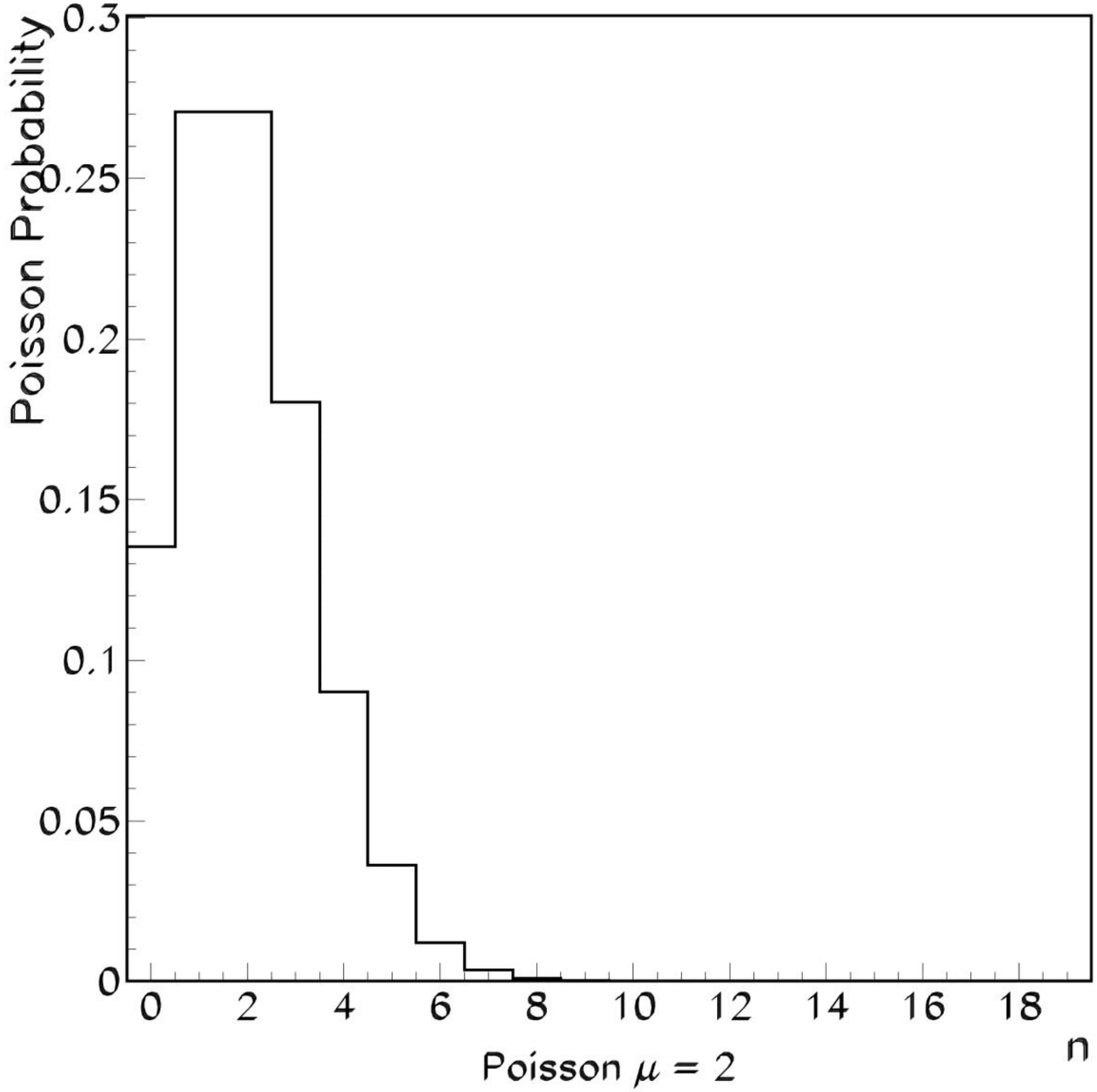
What Happens as μ Becomes Large?



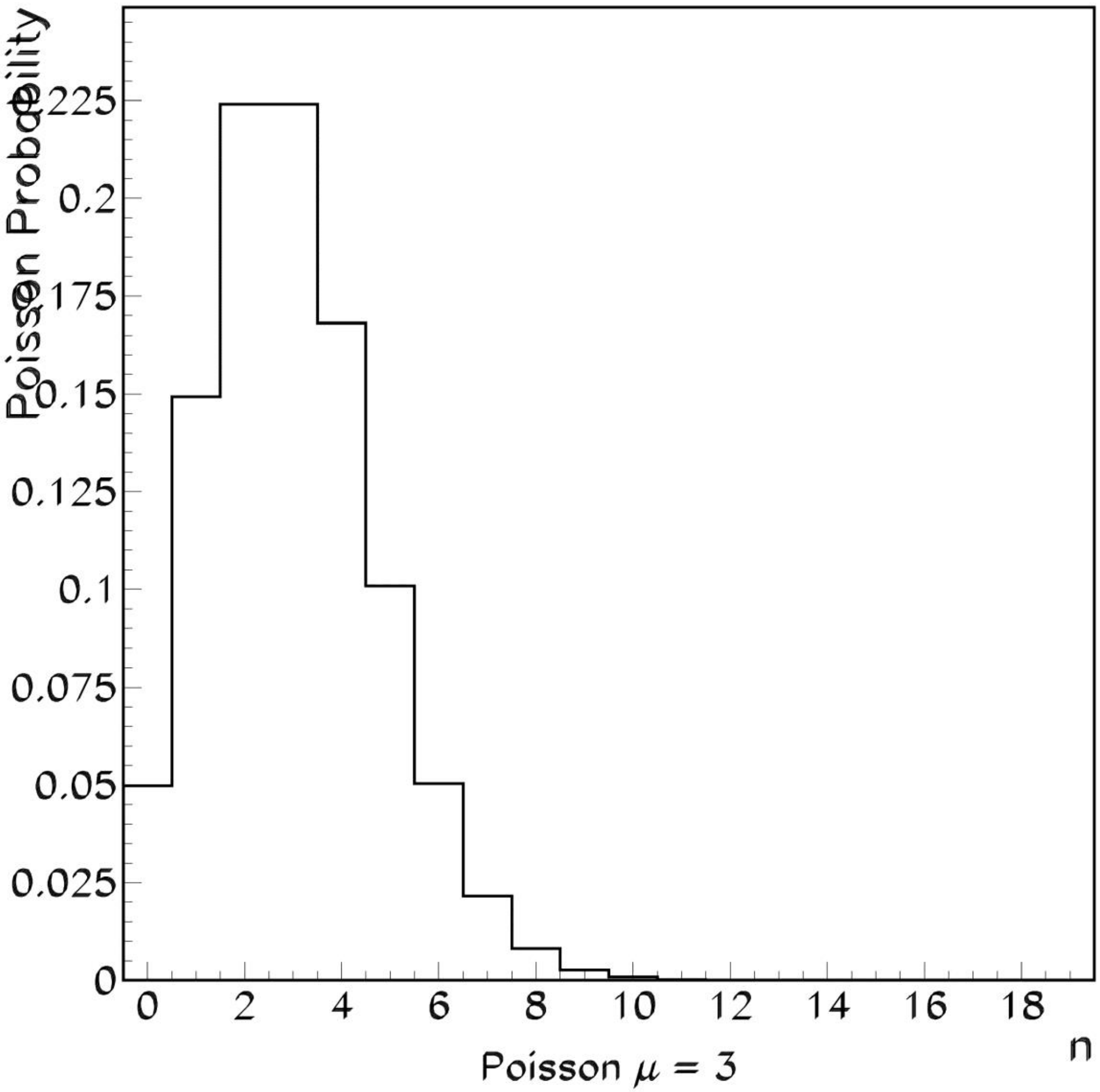
What Happens as μ Becomes Large?



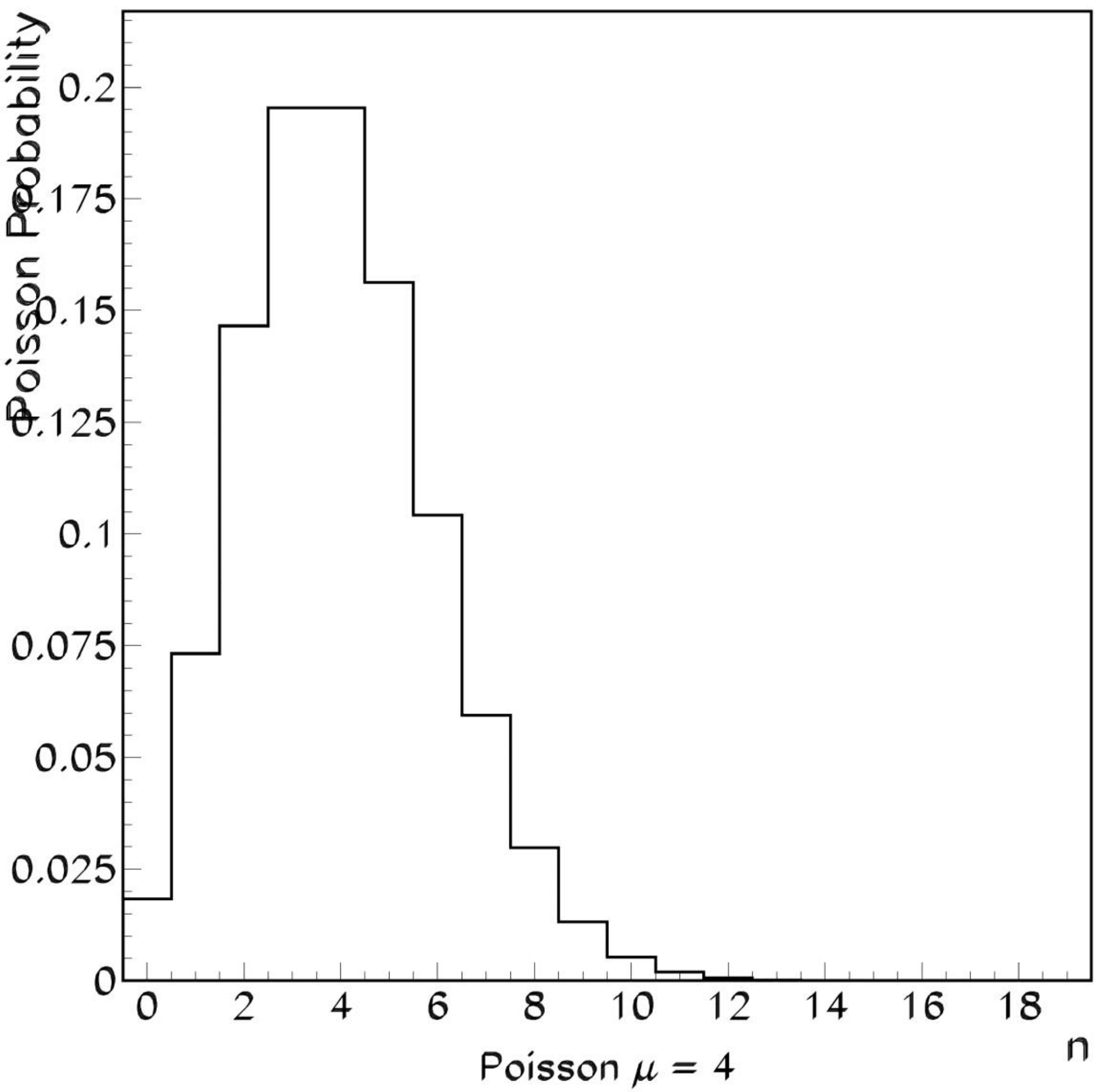
What Happens as μ Becomes Large?



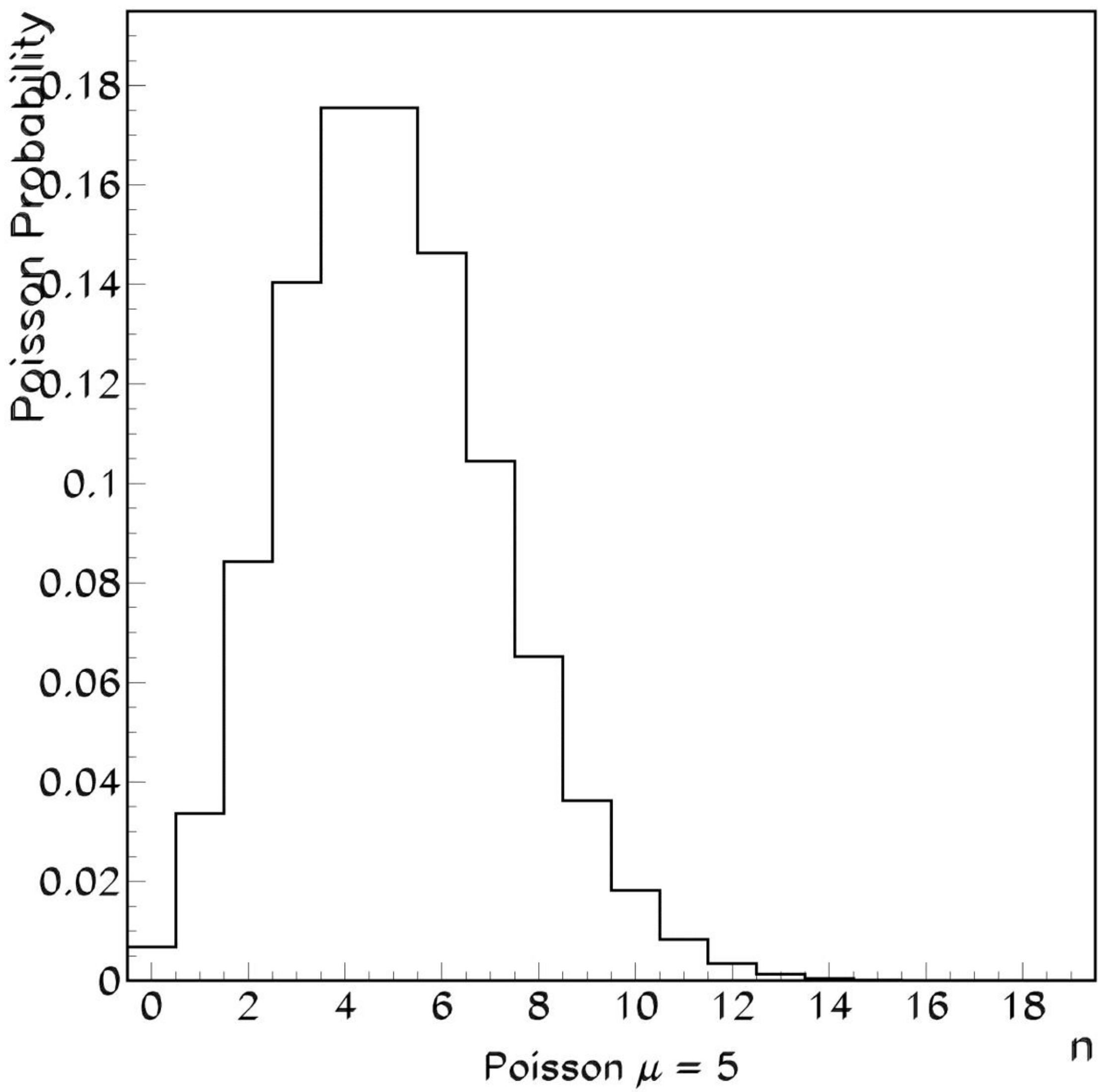
What Happens as μ Becomes Large?



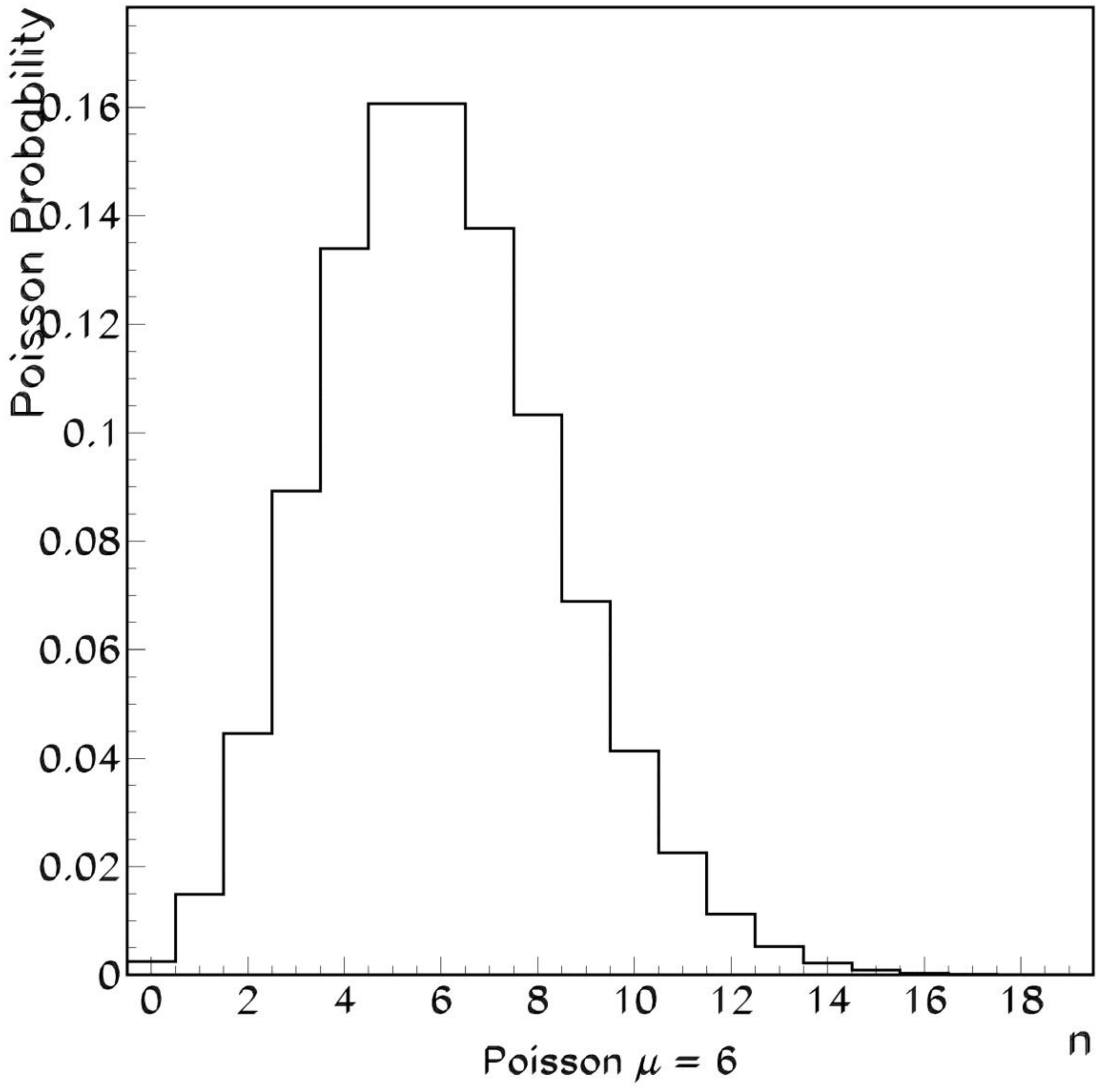
What Happens as μ Becomes Large?



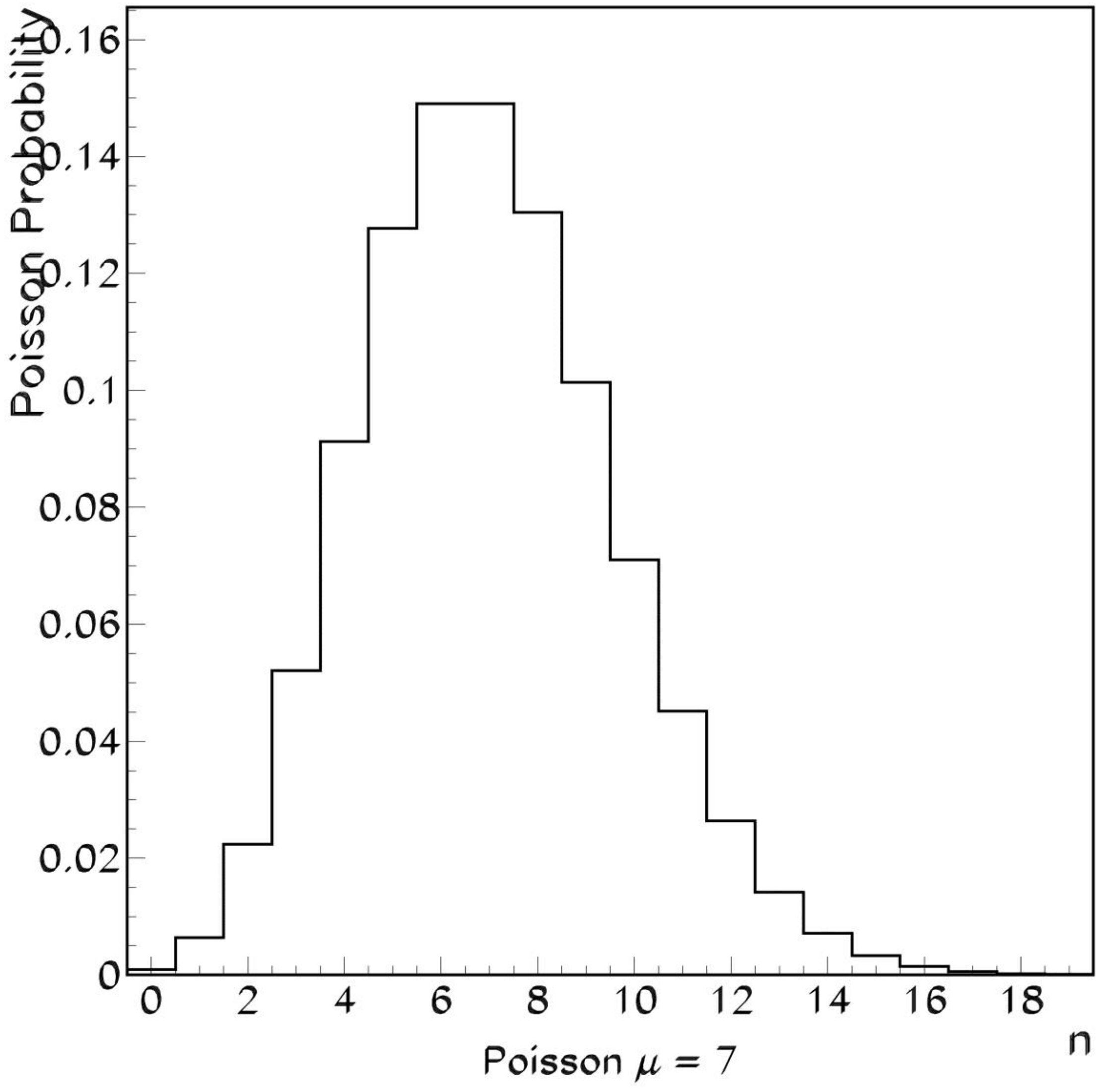
What Happens as μ Becomes Large?



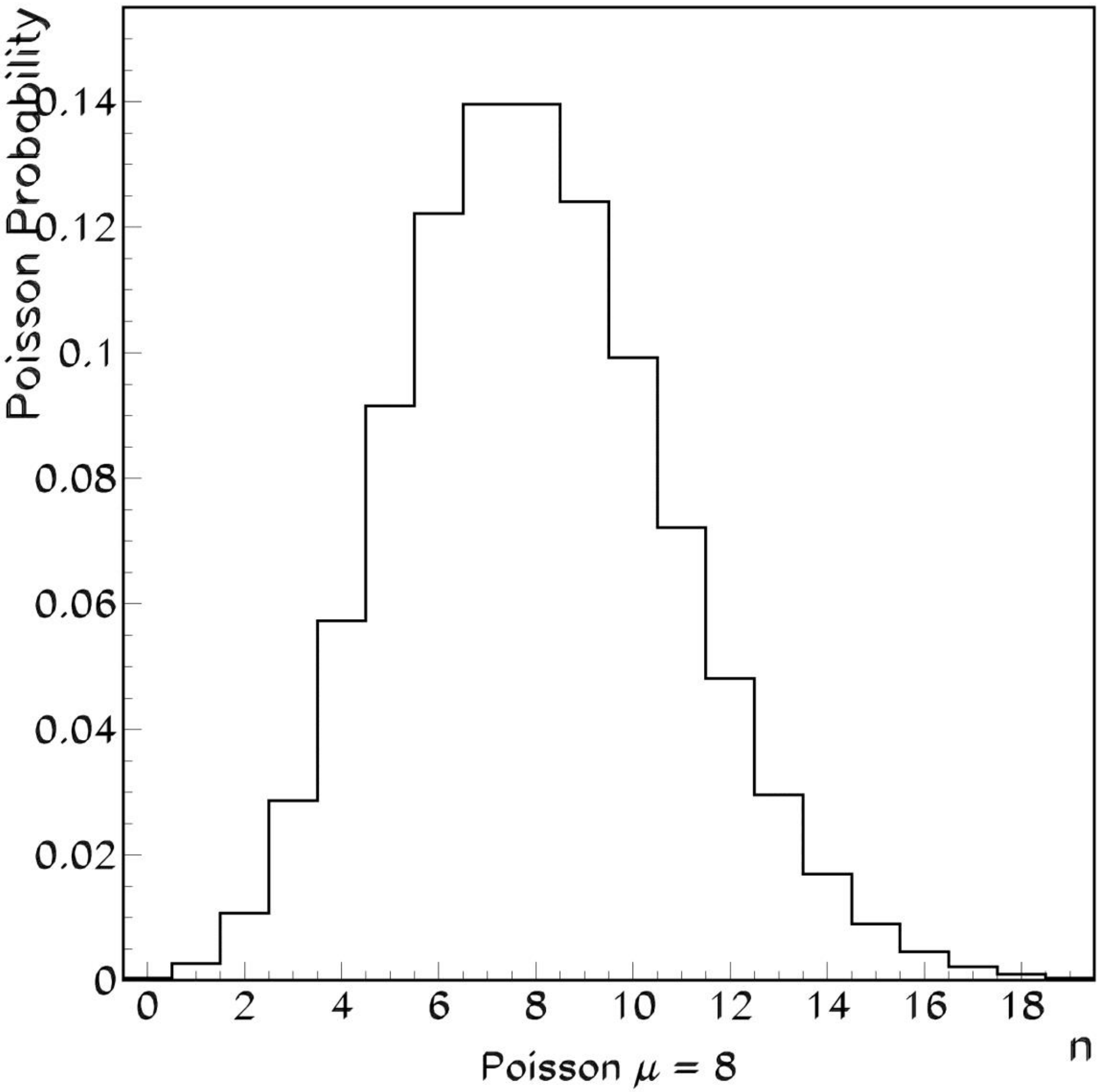
What Happens as μ Becomes Large?



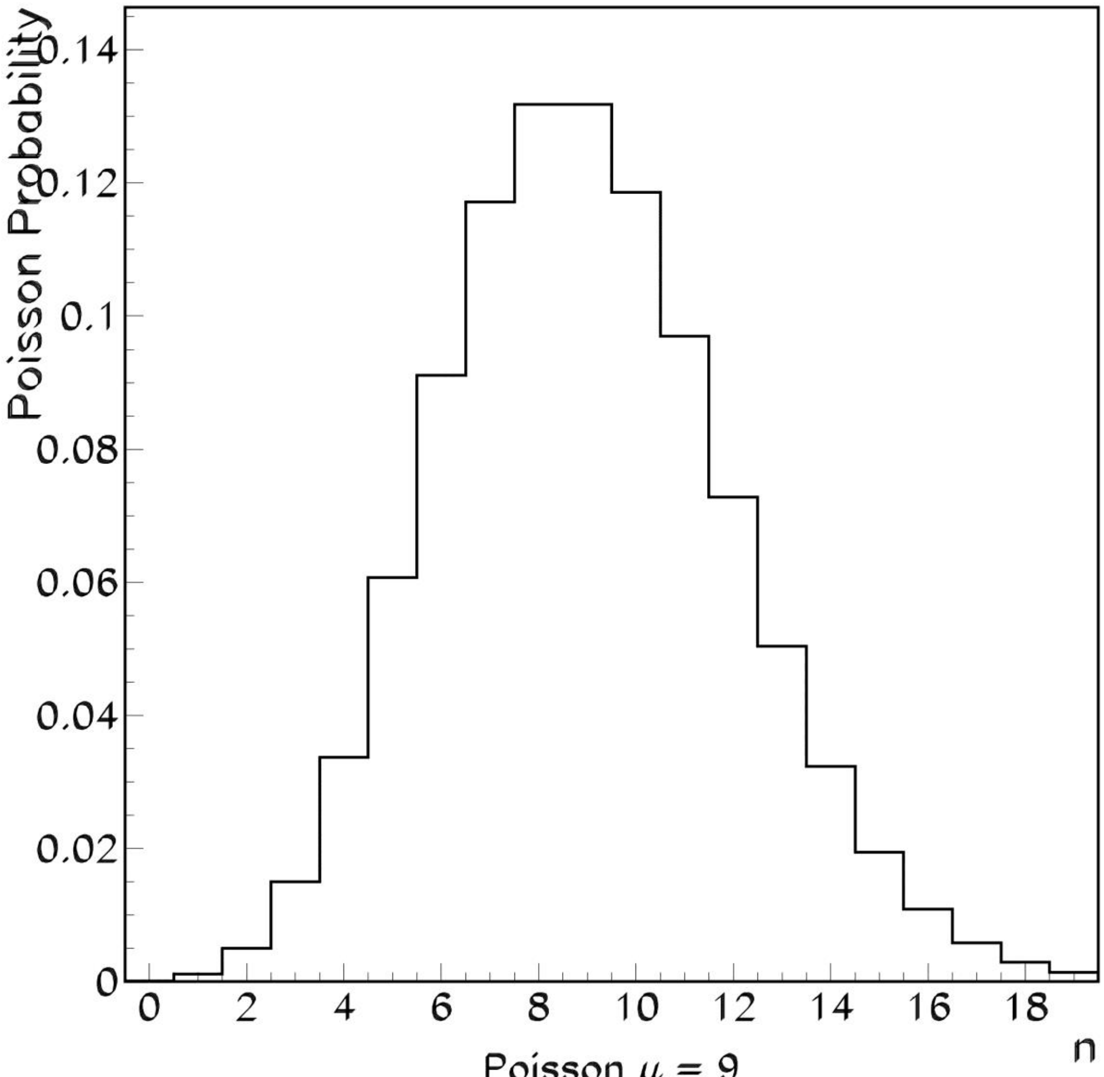
What Happens as μ Becomes Large?



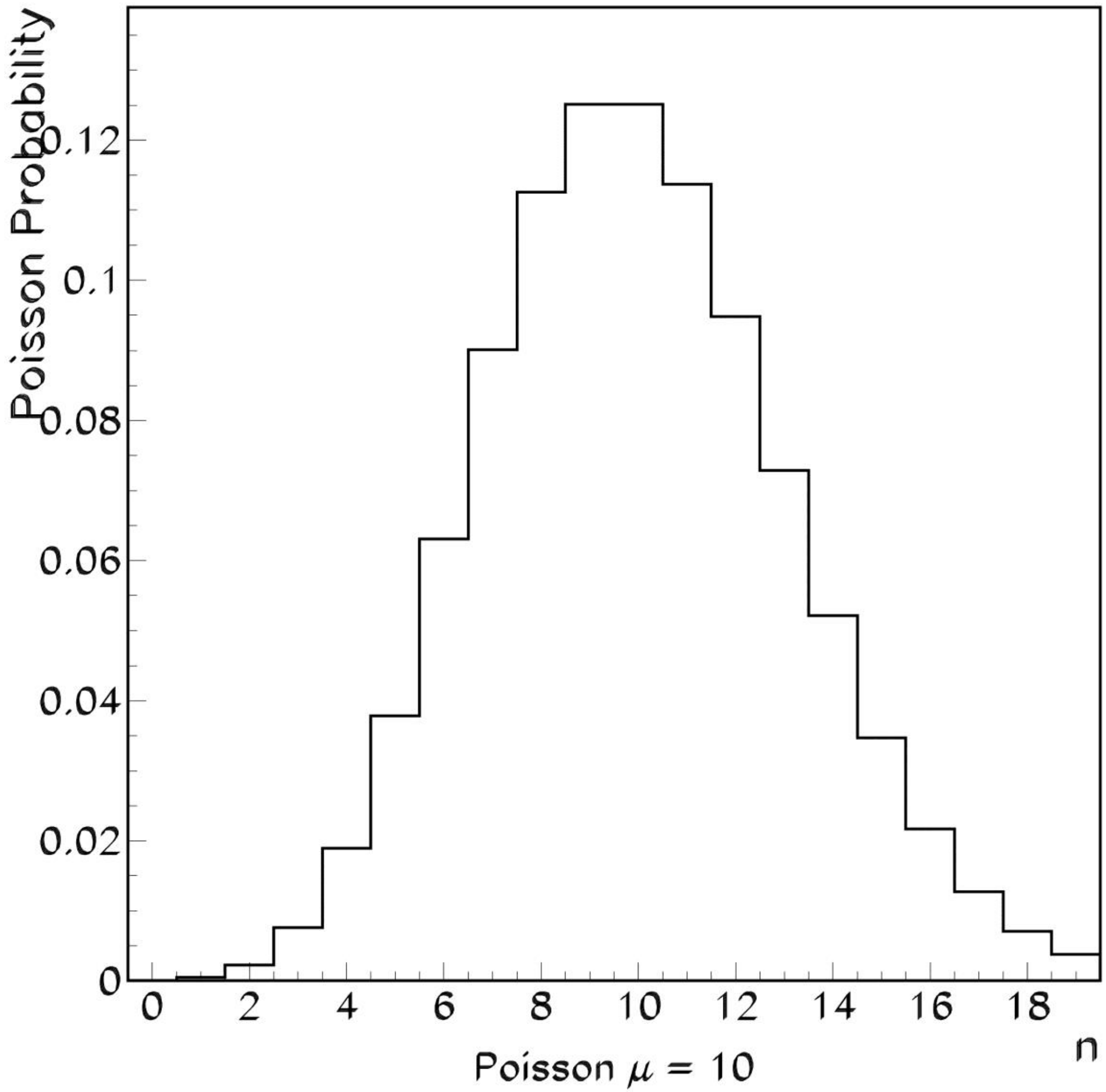
What Happens as μ Becomes Large?



What Happens as μ Becomes Large?



What Happens as μ Becomes Large?



Binomial Distribution

$$\begin{aligned} p &\longrightarrow 0 \\ N &\longrightarrow \infty \\ \text{S.T.} \quad \overline{\nu} = \mu &= \text{finite, small} \end{aligned}$$

Poisson Distribution

$$\begin{aligned} p &\longrightarrow 0 \\ N &\longrightarrow \infty \\ \text{S.T.} \quad \overline{\nu} = \mu &= \text{finite, large} \end{aligned}$$

Gaussian (Normal) Distribution

Gaussian (or Normal) Distribution: The probability P_G of observing n in a normally-distributed data set with mean μ is given by:

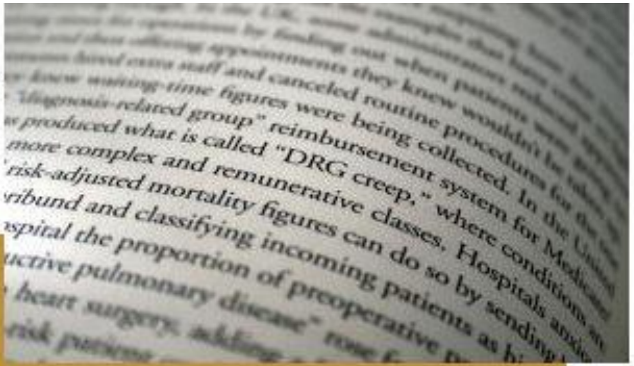
$$P_G(n;\mu) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-(n-\mu)^2/2\sigma_n^2}$$

The standard deviation of the distribution is given by

$$\sigma_n = \sqrt{\mu}$$

Finding Poisson Probabilities

Let X equal the number of typos on a printed page with a mean of 3 typos per page. What is the probability that a randomly selected page has **at least one typo** on it?



Solution. We can find the requested probability directly from the p.m.f. The probability that X is at least one is:

$$P(X \geq 1) = 1 - P(X = 0)$$

Therefore, using the p.m.f. to find $P(X = 0)$, we get:

$$P(X \geq 1) = 1 - \frac{e^{-3}3^{0}}{0!} = 1 - e^{-3} = 1 - 0.0498 = 0.9502$$

That is, there is just over a 95% chance of finding at least one typo on a randomly selected page when the average number of typos per page is 3.