Families of Random Variables

Fall 2017

Instructor:

Ajit Rajwade

Topic Overview

- We will study several useful families of random variables that arise in interesting scenarios in statistics.

- Discrete random variables- Bernoulli, Binomial, Poisson, multinomial, hypergeometric

- Continuous random variables- Gaussian, uniform, exponential, chi-square

Bernoulli Distribution

Definition

- Let X be a random variable whose value is 1 when a coin toss produces a heads, and a 0 otherwise. If p is the probability that the coin toss produces a heads, we have:

  $P(X = 1) = p$, $P(X = 0) = 1-p$

- This is called a **Bernoulli pmf** with parameter p – named after Jacob Bernoulli. X is called a Bernoulli random variable.

- Note here: the coin need not be unbiased any longer!

Properties

- E[X] = $p(1) + (1-p)(0) = p$.
- Var(X) = $p(1-p)^2 + (1-p)(0-p)^2 = p(1-p)$
- What's the median? $0.5 \text{ if } p = 0.5,\ 1 \text{ if } p > 0.5, \text{otherwise } 0$
- What is (are) the mode(s)? $\{0,1\} \text{ if } p = 0.5,\ 1 \text{ if } p > 0.5, \text{otherwise } 0$
- What is its MGF? $1 - p + pe^t$

Binomial Distribution

Definition

- Let *X* be a random variable denoting the number of heads in a sequence of *n* independent coin tosses (or *Bernoulli trials*) having *success probability* (i.e. probability of getting a heads) *p*.

- Then the pmf of *X* is given as follows:

$$P(X=i)=C(n,i)p^i(1-p)^{n-i}$$

- This is called the **binomial pmf** with parameters (*n*,*p*).

Defintion

- The pmf of X is given as follows:

$$P(X=i)=C(n,i)\,p^{i}(1-p)^{n-i}$$

$$C(n,i)=\frac{n!}{i!(n-i)!}$$

- Explanation: Consider a sequence of trials with i successes and n-i failures. The probability that this sequence occurs is $p^i(1-p)^{n-i}$, by the *product rule*. But there are C(n,i) such sequences – so we add their individual probabilities using the *sum rule*.

Definition

- Example: In 5 coin tosses, if we had two heads and three tails, the possible sequences are:

 HHTTT, HTHTT, HTTHT, HTTTH, TTTHH,
THTTH, TTHHT, TTTHH, THTHT,THHTT

- What's the probability that a sequence of Bernoulli trials produces a success only on the *i*-th trial? Note that this is **not** a binomial distribution.

Definition

- The pmf of X is given as follows:

$$P(X = i) = C(n,i)\,p^i(1-p)^{n-i}$$

$$C(n,i) = \frac{n!}{i!(n-i)!}$$

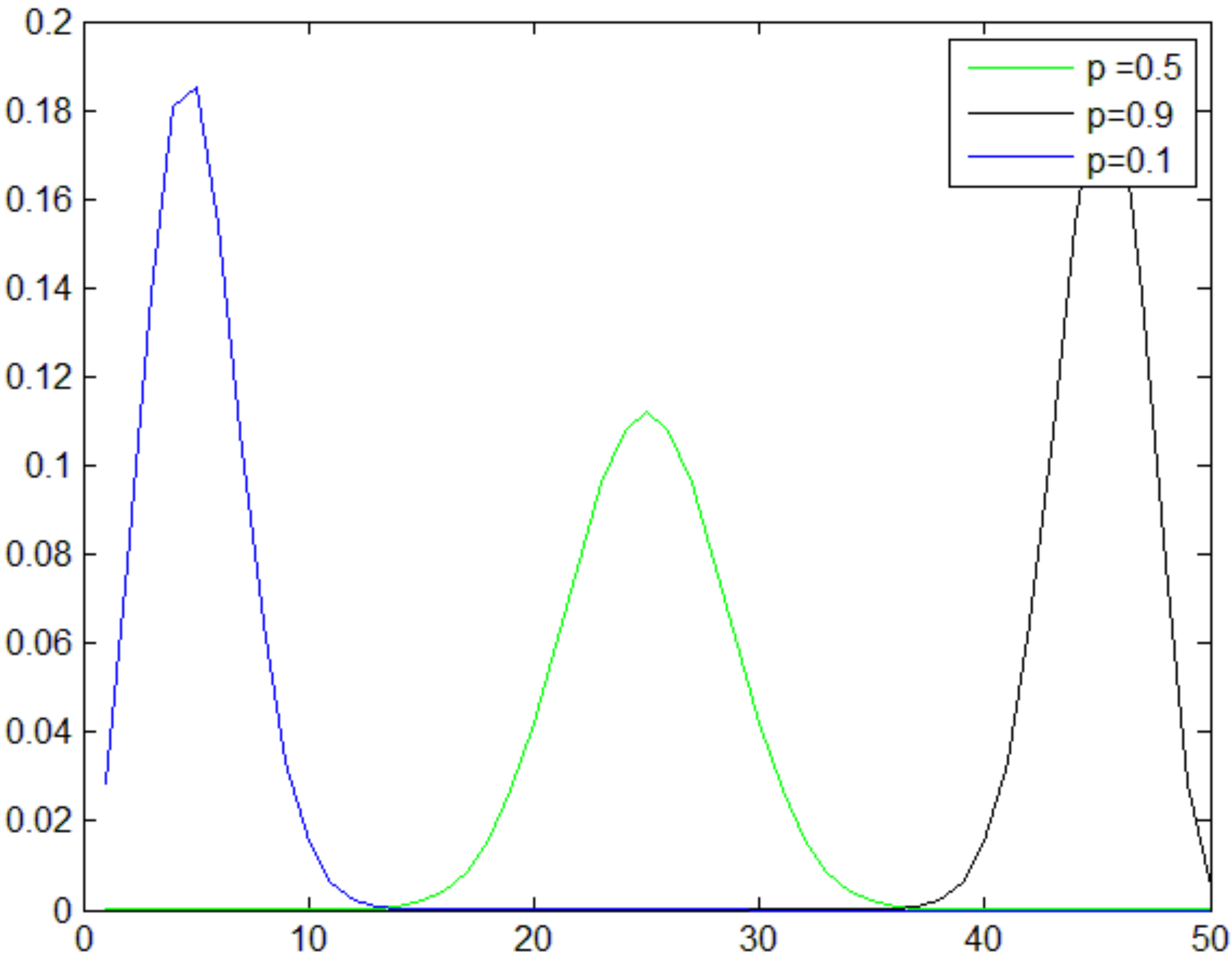
- To verify it is a valid pmf:

$$\sum_{i=1}^{n} P(X = i) = \sum_{i=1}^{n} C(n,i)\,p^i(1-p)^{n-i}$$

$$= (p + (1-p))^n$$

$$= 1$$

Binomial theorem!


0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
10
20
30
40
50
p =0.5
p=0.9
p=0.1

Example 1

- Let's say you design a smart self-driving car and you have tested it thoroughly. You have determined that the probability that your car collides is p. You go for an international competition and the rules say that you will win a prize if in k difficult tests, your car collides at most once. What's the probability that you will win the prize?

- Answer: X is the number of times your car collides. X is a binomial random variable with parameters (k,p). The probability that you win a prize is

  $P(X <= 1) = (1-p)^k + C(k,1)p(1-p)^{k-1}$.

Example 2

- At least half of an airplane's engines need to function for it to operate. If each engine independently fails with probability p, for what values of p is a 4-engine airplane more likely to operate than a 2-engine airplane?
- Answer: The number of functioning engines (X) is a binomial random variable with parameters $(4,p)$ or $(2, p)$.
- In 4-engine case, the probability of operation is $P(X=2) + P(X=3) + P(X=4) = 6p^2(1-p)^2 + 4p(1-p)^3+(1-p)^4$.
- In 2-engine case, it is $P(X=1)+P(X=2) = 2p(1-p)+(1-p)^2$. Do the math!
- Answer is for $p < 1/3$.

Properties

- Mean: Recall that binomial random variable *X* is the *sum* of random variables of the following form:

$$X_i = 1 \text{ if trial } i \text{ yields success (this occurs with prob. } p)$$
$$= 0 \text{ otherwise}$$

- Hence

$$E(X) = E(\sum_{i=1}^{n} X_i) = np$$

Notice how we are making use of the linearity of the expectation operator. These calculations would have been much harder had you tried to plug in the formulae for binomial distribution.

- Variance:

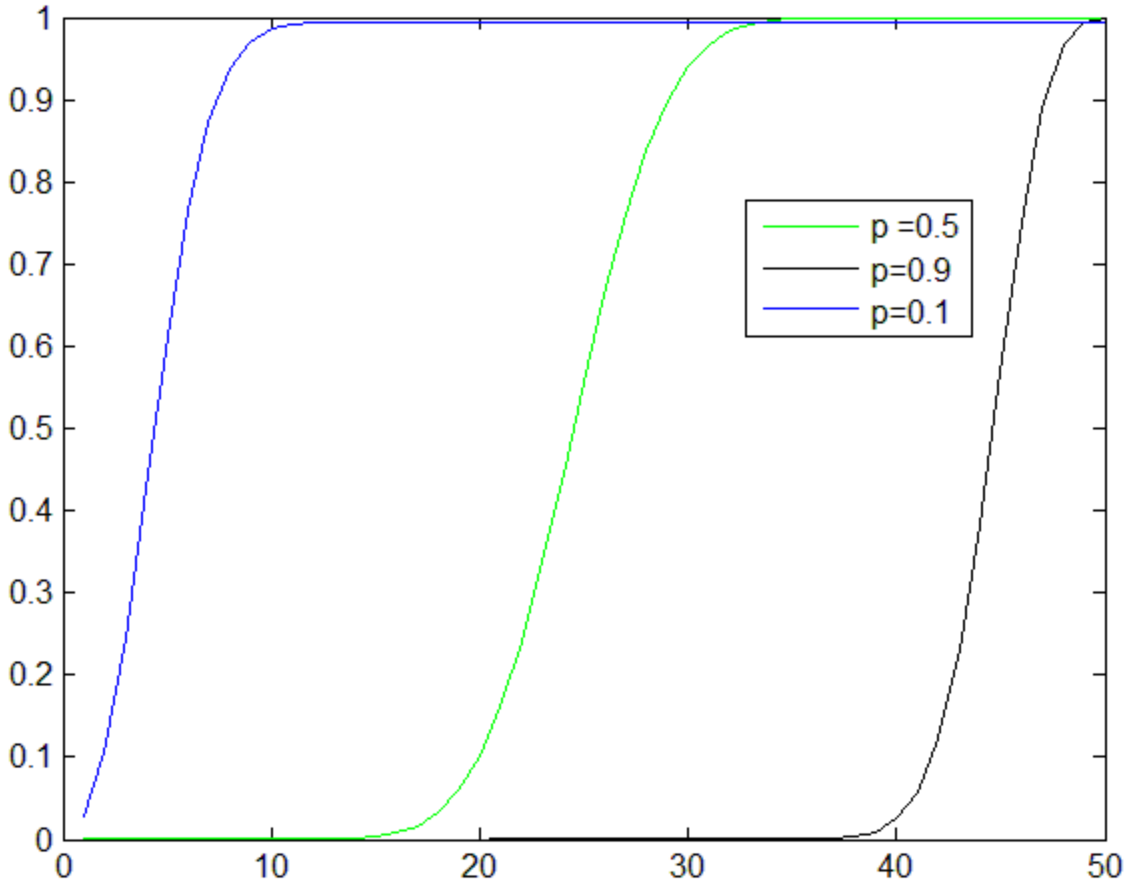
$$Var(X) = Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) = np(1-p)$$

Properties

- The CDF is given as follows:

$$F_X(i) = P(X \le i) = \sum_{k=0}^{i} C(n,k)\, p^k (1-p)^{n-k}$$

- MGF: Consider a binomial r.v. to be the sum of *n* independent Bernoulli trials. Hence using a property of the MGF (which one?), the MGF of the binomial r.v. is given by $(1-p+pe^t)^n$.


p =0.5
p=0.9
p=0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50

Example

Bits are sent over a communications channel in packets of 12. If the probability of a bit being corrupted over this channel is 0.1 and such errors are independent, what is the probability that no more than 2 bits in a packet are corrupted?

If 6 packets are sent over the channel, what is the probability that at least one packet will contain 3 or more corrupted bits?

Let X denote the number of packets containing 3 or more corrupted bits. What is the probability that X will exceed its mean by more than 2 standard deviations?

Example

- We want $P(X = 0) + P(X=1) + P(X=2)$.
- $P(X=0) = (0.1)^0(0.9)^{12} = 0.282$
- $P(X=1) = C(12,1)\ (0.1)^1(0.9)^{11} = 0.377$
- $P(X=2) = C(12,2)\ (0.1)^2(0.9)^{10}=0.23$
- So the answer is 0.889.

Example

- The probability that 3 or more bits are corrupted is 1-0.889 = 0.111
- Let Y = number of packets with 3 or more corrupted bits. Then we want $P(Y \geq 1) = 1-P(Y=0)$
  $= 1-C(6,0)(0.111)^0 (0.889)^6 = 0.4936$.
- Mean of Y is $\mu = 6(0.111) = 0.666$.
- Standard deviation of Y is $\sigma = [6(0.111)(0.889)]^{0.5} = 0.77$.
- We want $P(Y > \mu + 2\sigma) = P(Y > 2.2) = P(Y \geq 3) = 1-P(Y=0)-P(Y=1)-P(Y=2) = ?$

Example

- We want $P(Y > \mu + 2\sigma) = P(Y > 2.2) = P(Y \geq 3) = 1 - P(Y=0) - P(Y=1) - P(Y=2) = ?$

- $P(Y=0)=C(6,0)\ (0.111)^0(0.889)^6 = 0.4936$
- $P(Y=1) = C(6,1)(0.111)(0.889)^5 = 0.37$
- $P(Y=2) = C(6,2)\ (0.111)^2(0.889)^4 = 0.115$

- $P(Y > \mu + 2\sigma) = 1-(0.4936+0.37+0.115) = 0.0214$

Related distributions

- In a sequence of Bernoulli trials, let *X* be the random variable for the trial number that gave the first success.

- Then *X* is called a geometric random variable and its pmf is given as:

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

- Let *X* be the trial number for the *k*th success in a sequence of Bernoulli trials. Then *X* is called a **negative binomial random variable**. What is its pmf?

$$P(X = k) = C(n-1, k-1)\,p^k(1-p)^{n-k}$$

Properties: Mode

- Let $X \sim$ Binomial (n,p).

- Then $P(X = k) \geq P(X = k\text{-}1) \leftrightarrow k \leq (n+1)p$ (prove this).

- Also $P(X = k) \geq P(X = k+1) \leftrightarrow k \geq (n+1)p\text{-}1$ (prove this).

- Any integer-valued k which satisfies both the above conditions will be the mode (why?).

- If $(n+1)p$ is an integer then that's the mode of the binomial. Otherwise the binomial has two modes – and they are consecutive integers.

Poisson distribution

Definition – and genesis

- We have seen the binomial distribution before:

$$P(X=i)=\frac{n!}{i!(n-i)!}p^i(1-p)^{n-i}$$

- Here *p* is the success probability. We can express it in the form

$$p=\frac{\lambda}{n}, \lambda = \text{expected number of successes in } n \text{ trials}$$

- Hence

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

Definition – and genesis

- We have

$$P(X=i)=\frac{n(n-1)(n-2)\ldots(n-i+1)}{i!}\left(\frac{\lambda}{n}\right)^{i}(1-\frac{\lambda}{n})^{n-i}$$

- In the limit when $n \to \infty$ and $p \to 0$ such that $\lambda = np$, we have

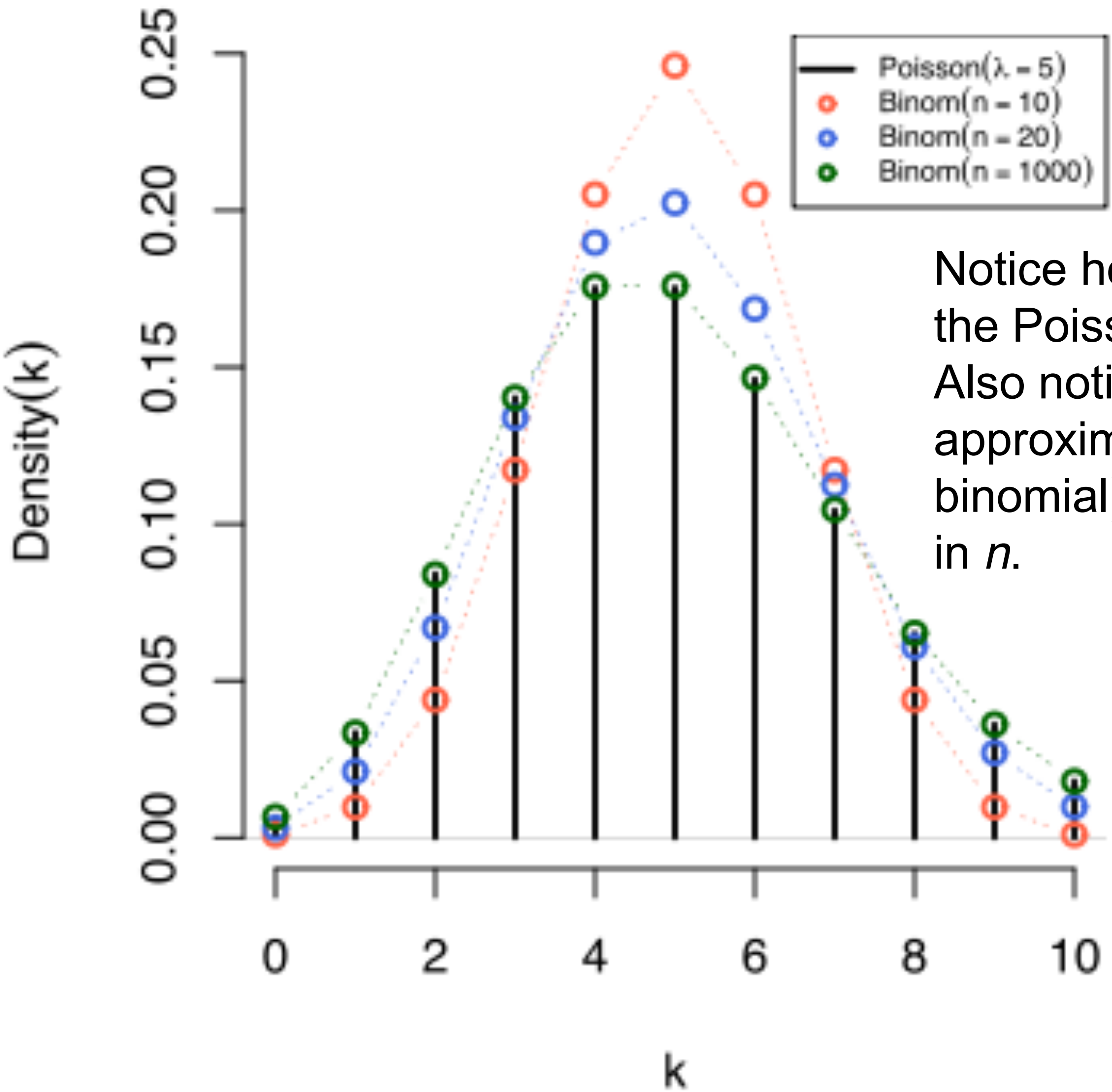
$$P(X=i)=\frac{\lambda^{i}}{i!}e^{-\lambda}$$

- This is called as the **Poisson pmf** and the above statement is called the Poisson limit theorem.

Definition

- The Poisson distribution is used to model the number of successes of a long sequence of independent Bernoulli trials if the *expected number of successes* (i.e. λ) is *known* and *constant*.

- For a Poisson random variable, note that the expected number of successes λ is constant and the **parameter** of the pmf. This is unlike the Binomial pmf for which the success probability p of a *single* Bernoulli trial is constant and also a parameter of the pmf.



Notice how the binomial is resembling the Poisson with an increase in *n*. Also notice that *np* (which is approximately the mode of the binomial) is constant despite increase in *n*.

Properties

- To double check that it is indeed a valid pmf, we check that:

$$\sum_{i=0}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1$$

Using Taylor's series for exponential function about 0

- The afore-mentioned analysis tells us that the expected number of successes is equal to λ. To prove this rigorously – see next slide.

Properties

- The afore-mentioned analysis tells us that the expected number of successes is equal to λ. To prove this rigorously:

$$E(X) = \left(\sum_{i=0}^{\infty} i e^{-\lambda} \frac{\lambda^i}{i!} \right)$$

$$= \left(\sum_{i=1}^{\infty} i e^{-\lambda} \frac{\lambda^i}{i!} \right)$$

$$= \lambda e^{-\lambda} \left(\sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \right)$$

$$= \lambda e^{-\lambda} \left(\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \right)$$

$$= \lambda e^{-\lambda} e^{\lambda} = \lambda$$

Properties

- Variance:

$$Var(X) = E(X^2) - (E(X))^2$$
$$E(X^2) = \left(\sum_{i=0}^{\infty} i^2 e^{-\lambda} \frac{\lambda^i}{i!}\right) = \lambda^2 + \lambda$$
$$\therefore Var(X) = \lambda$$

Detailed proof on the board. Also see here.

- MGF:

$$\Phi_X(t) = E(e^{tX})$$
$$= \sum_{i=0}^{\infty} e^{ti} e^{-\lambda} \lambda^i / i! = e^{-\lambda} \sum_{i=0}^{\infty} (\lambda e^t)^i / i!$$
$$= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

Properties

- Mode:

k is a mode if

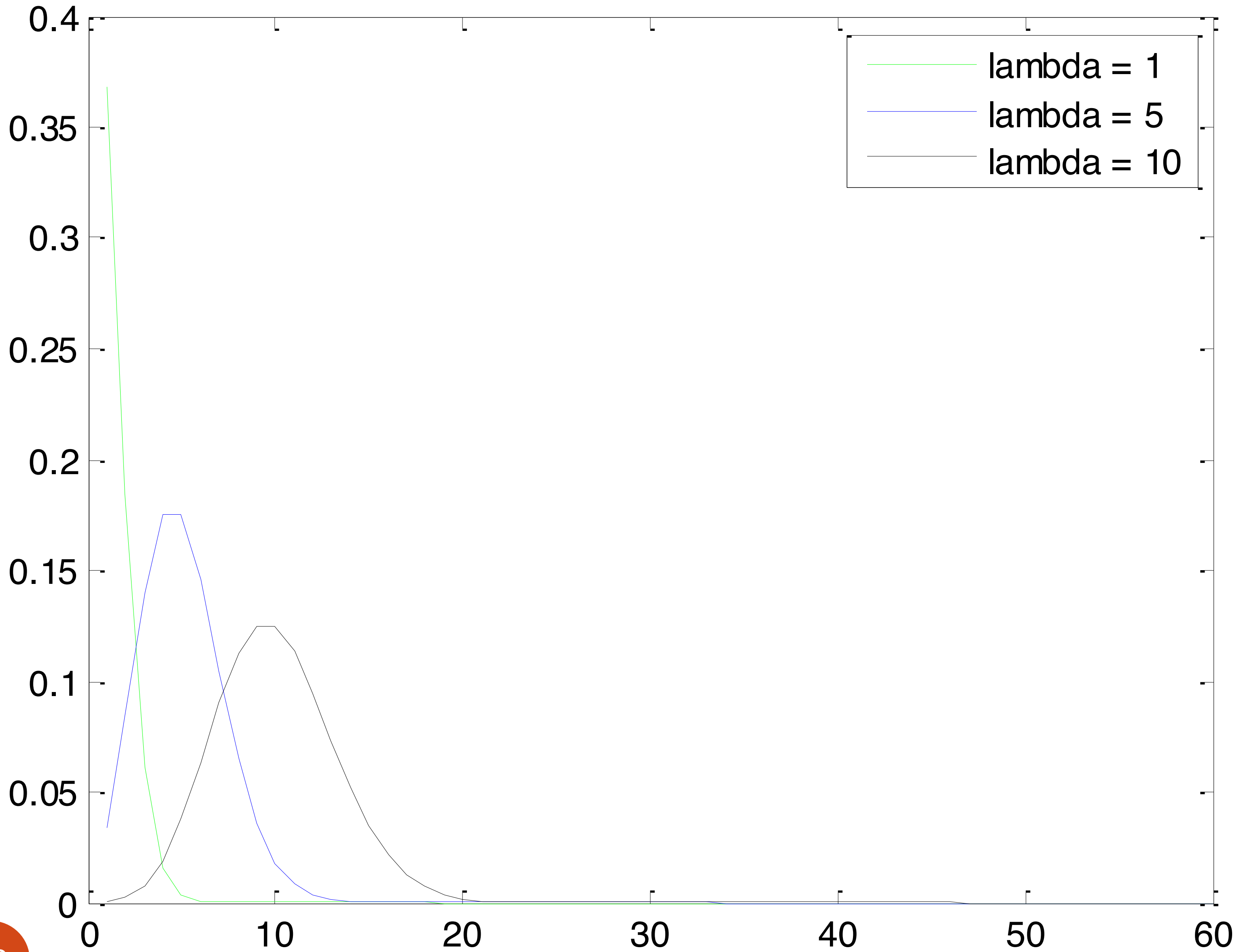
$$P(X=k) \geq P(X=k+1) \rightarrow k \geq \lambda - 1$$

and

$$P(X=k) \geq P(X=k-1) \rightarrow k \leq \lambda$$

Thus we seek an **integer** k that satisfies both these conditions - note that often, λ is **not** an integer.

Notice: the mean and variance both increase with lambda.



Properties

- Consider independent Poisson random variables *X* and *Y* having parameters λ_1 and λ_2 respectively. Then $Z = X + Y$ is also a Poisson random variable with parameter $\lambda_1 + \lambda_2$.

  Detailed proof on the board and in tutorial 1.

- PMF – recurrence relation:

$$P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}$$

$$P(X = i+1) = \frac{\lambda^{i+1}}{(i+1)!} e^{-\lambda}$$

$$\frac{P(X = i+1)}{P(X = i)} = \frac{\lambda}{i+1}, P(X = 0) = e^{-\lambda}$$

Properties

- If $X \sim$ Poisson (λ), $P(Y|X=l)$ = Binomial(l,p) where $\lambda > 0$ and $0 \leq p \leq 1$, then $Y \sim$ Poisson (λp). This is called as **thinning of a Poisson random variable by a Binomial**. *We will cover this derivation in a tutorial.*

Poisson distribution: examples

- The number of misprints in a book (assuming the probability p of a misprint is small, and the number n of letters typed is very large, with np = expected number of misprints remaining constant)

- Number of traffic rule violations in a typical city in the USA (assuming the probability p of a violation is small, and the number of vehicles is very large).

- In general, the Poisson distribution is used to model rare events, even though the event has plenty of "opportunities" to occur. (Sometimes called the **law of rare events** or the **law of small numbers**).

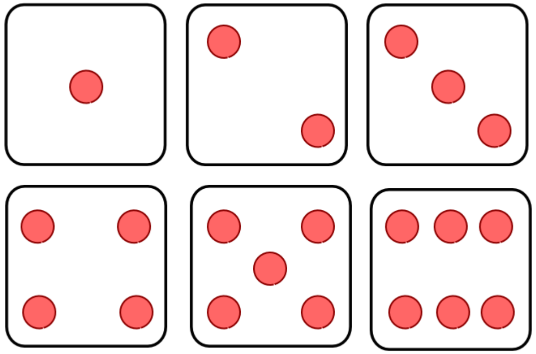
Poisson distribution: examples

- Number of people in a country who live up to 100 years
- Number of wrong numbers dialed in a day
- Number of laptops that fail on the first day of use
- Number of photons of light counted by the detector element of a camera

Multinomial distribution

Definition

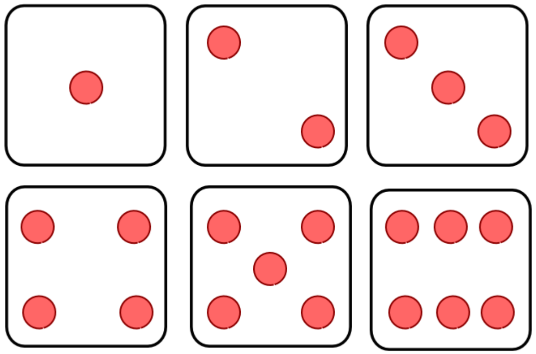
- Consider a sequence of n independent trials each of which will produce one out of k possible outcomes, where the set of possible outcomes is the *same* for each trial.



- Assume that the probability of each of the k outcomes is known and constant and given by p_1, p_2, …, p_k.

Definition

- Let **X** be a k-dimensional random variable for which the i^{th} element represents the number of trials that produced the i^{th} outcome (also known as the number of *successes* for the i^{th} category)



Eg: in 20 throws of a die, you had 2 ones, 4 twos, 7 threes, 4 fours, 1 five and 2 sixes.

Definition

- Then the pmf of **X** is given as follows:

$$
\begin{aligned}
&P(\mathbf{X} = (x_1, x_2, \ldots, x_k)) \\
&= P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) \\
&= \frac{n!}{x_1!\,x_2!\ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} \\
&\forall i, 0 \le p_i \le 1, \sum_{i=1}^{n} p_i = 1, x_1 + x_2 + \ldots + x_k = n
\end{aligned}
$$

The number of ways to arrange *n* objects which can be divided into *k* groups of identical objects. There are x_1 objects of type 1, x_2 objects of type 2, and x_k objects of type *k*.

- This is called the **multinomial pmf**.

Definition

- The success probabilities for each category, i.e. p_1, p_2,…, p_k are all *parameters* of the multinomial pmf.

- Remember: The multinomial random variable is a vector whose i^{th} component is the number of successes of the i^{th} category (i.e. the number of times that the trials produced a result of the i^{th} category).

Properties

- Mean ***vector***:

$$E(\mathbf{X}) = (np_1, np_2, \ldots, np_k), E(X_i) = np_i$$

- Variance of a component

$$Var(X_i) = Var(\sum_{j=1}^{n} X_{ij}) = \sum_{j=1}^{n} Var(X_{ij}) = np_i(1 - p_i)$$

Assuming independent trials

X_{ij} is a Bernoulli random variable which tells you whether or not there was a success in the *i*th category on the *j*th trial

Properties

- For vector-valued random variables, the variance is replaced by the **covariance matrix**. The covariance matrix **C** in this case will have size *k* x *k*, where we have:

$$C(i,j) = E[(X_i - \mu_i)(X_j - \mu_j)] = Cov(X_i, X_j)$$

$$Cov(X_i, X_j) = -np_i p_j, i \neq j$$

$$= np_i(1 - p_i), i = j$$

$$\text{Proof : next page}$$

$$Cov(X_i, X_j) = -np_i p_j$$

Proof :

$$X_i = \#\text{successes in category } i = \sum_{k=1}^{n} X_{ik}$$

$$X_j = \#\text{successes in category } j = \sum_{l=1}^{n} X_{jl}$$

These are independent Bernoulli random variables – each representing the outcome of a trial (indexed by *k* and *l*)

$$Cov(X_i, X_j) = \sum_{k=1}^{n}\sum_{l=1}^{n} Cov(X_{ik}, X_{jl})$$

By linearity of covariance

$$= \sum_{k=1}^{n}\sum_{l=1, l\neq k}^{n} Cov(X_{ik}, X_{jl}) + \sum_{l=1}^{n} Cov(X_{il}, X_{jl})$$

By independence of trials

$$= 0 + \sum_{l=1}^{n} (E(X_{il}X_{jl}) - E(X_{il})E(X_{jl}))$$

Since in a trial, success can be achieved only in one category

$$= \sum_{l=1}^{n} (0 - p_i p_j) = -np_i p_j$$

MGF for a Multinomial

- For k = 2, the multinomial reduces to the binomial.
- Let us derive the MGF for k = 3 (trinomial):

$$\boldsymbol{X} = (X_1, X_2), \boldsymbol{x} = (x_1, x_2)$$

$$P(\boldsymbol{X} = \boldsymbol{x}) = \frac{n!}{x_1!x_2!(n - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{n - x_1 - x_2}$$

$$\Phi_X(\boldsymbol{t}) = \Phi_X(t_1, t_2) = E(e^{t_1 X_1 + t_2 X_2})$$

$$= \sum_{x_1=0}^{n} \sum_{x_2=0}^{n-x_1} \frac{n!}{x_1!x_2!(n - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{n - x_1 - x_2} e^{t_1 x_1} e^{t_2 x_2}$$

$$= \sum_{x_1=0}^{n} \sum_{x_2=0}^{n-x_1} \frac{n!}{x_1!x_2!(n - x_1 - x_2)!} \left(p_1 e^{t_1}\right)^{x_1} \left(p_2 e^{t_2}\right)^{x_2} (1 - p_1 - p_2)^{n - x_1 - x_2}$$

$$= (p_1 e^{t_1} + p_2 e^{t_2} + 1 - p_1 - p_2)^n$$

This follows from the multinomial theorem.

MGF for a Multinomial

- Multinomial theorem:

$$(x_1 + x_2 + \ldots + x_m)^n = \sum_{k_1+k_2+\ldots+k_m=n} \frac{n!}{k_1!k_2!\ldots k_m!} \prod_{i=1}^{m} x_i^{k_i}$$

- For arbitrary *k*:

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_{k-1}), \boldsymbol{x} = (x_1, x_2, \ldots, x_{k-1})$$

$$\Phi_X(\boldsymbol{t}) = (p_1 e^{t_1} + p_2 e^{t_2} + \ldots + p_{k-1} e^{t_{k-1}} + 1 - p_1 - p_2 - \ldots - p_{k-1})^n$$

Hypergeometric distribution

Sampling with and without replacement

- Suppose there are k objects each of a different type.

- When you sample 2 objects from these **with replacement**, you pick a particular object with probability $1/k$, and you place it back (***replace*** it).

- The probability of picking an object of another type is again $1/k$.

- When you sample **without replacement**, the probability that your first object was of so and so type is $1/k$. The probability that your second object was of so and so type is now $1/(k-1)$ because you ***didn't*** put the first object back!

Definition

- Consider a set of objects of which N are of good quality and M are defective.
- Suppose you pick some n objects out of these *without* replacement.
- There are $C(N+M,n)$ ways of doing this.
- Let X be a random variable denoting the number of good quality objects picked (out of a total of n).

Definition

- There are C(N,i)C(M,n-i) ways to pick i good quality objects and n-i bad objects.

- So we have

$$P(X = i) = \frac{C(N,i)C(M,n-i)}{C(N+M,n)}, 0 \leq i \leq n$$

$$C(a,b) = 0 \text{ if } b > a \text{ or } b < 0$$

- Such a random variable X is called a **hypergeometric random variable**.

Properties

- Consider random variable X_i which has value 1 if the i^{th} trial produces a good quality object and 0 otherwise.

- Now consider the following probabilities:

$$P(X_1 = 1) = \frac{N}{N+M}$$

$$P(X_2 = 1) = P(X_2 = 1 \mid X_1 = 1)P(X_1 = 1) +$$

$$P(X_2 = 1 \mid X_1 = 0)P(X_1 = 0)$$

$$= \frac{N-1}{N+M-1}\frac{N}{N+M} + \frac{N}{N+M-1}\frac{M}{N+M} = \frac{N}{N+M}$$

In general, $P(X_i = 1) = \dfrac{N}{N+M}$

Properties

- Note that:

Each X_i is a Bernoulli random variable with parameter $p=N/(N+M)$.

$$X = \sum_{i=1}^{n} X_i$$

$$\therefore E(X) = E\left(\sum_{i=1}^{n} X_i\right) = \frac{nN}{N+M}$$

$$\therefore Var(X) = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} Cov(X_i, X_j)$$

$$Var(X_i) = P(X_i = 1)(1 - P(X_i = 1)) = \frac{NM}{N+M}$$

Properties

- Note that:

$$\therefore Var(X) = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} Cov(X_i, X_j)$$

$$Var(X_i) = P(X_i = 1)(1 - P(X_i = 1)) = \frac{NM}{N+M}$$

$$Cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$$

$$E(X_i X_j) = P(X_i X_j = 1) = P(X_j = 1, X_i = 1)$$

$$= P(X_j = 1 \mid X_i = 1) P(X_i = 1)$$

$$= \frac{N-1}{N+M-1}\frac{N}{N+M}$$

Properties

- Note that:

$$Cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$$

$$E(X_i X_j) = P(X_i X_j = 1) = P(X_j = 1, X_i = 1)$$

$$= P(X_j = 1 \mid X_i = 1)P(X_i = 1)$$

$$= \frac{N-1}{N+M-1}\frac{N}{N+M}$$

$$Cov(X_i, X_j) = \frac{N(N-1)}{(N+M)(N+M-1)} - \left(\frac{N}{N+M}\right)^2$$

$$= \frac{-NM}{(N+M)^2(N+M-1)}$$

Properties

- Note that:

$$Cov(X_i, X_j) = \frac{N(N-1)}{(N+M)(N+M-1)} - \left(\frac{N}{N+M}\right)^2$$

$$= \frac{-NM}{(N+M)^2(N+M-1)}$$

$$Var(X) = \frac{nNM}{(N+M)^2} - \frac{n(n-1)NM}{(N+M)^2(N+M-1)}$$

$$= \frac{nNM}{(N+M)^2}\left(1 - \frac{n-1}{N+M-1}\right)$$

$$= np(1-p)\left(1 - \frac{n-1}{N+M-1}\right)$$

Recall: Each X_i is a Bernoulli random variable with parameter $p=N/(N+M)$.

$\approx np(1-p)$ when N and/or M is/are very large

Gaussian distribution

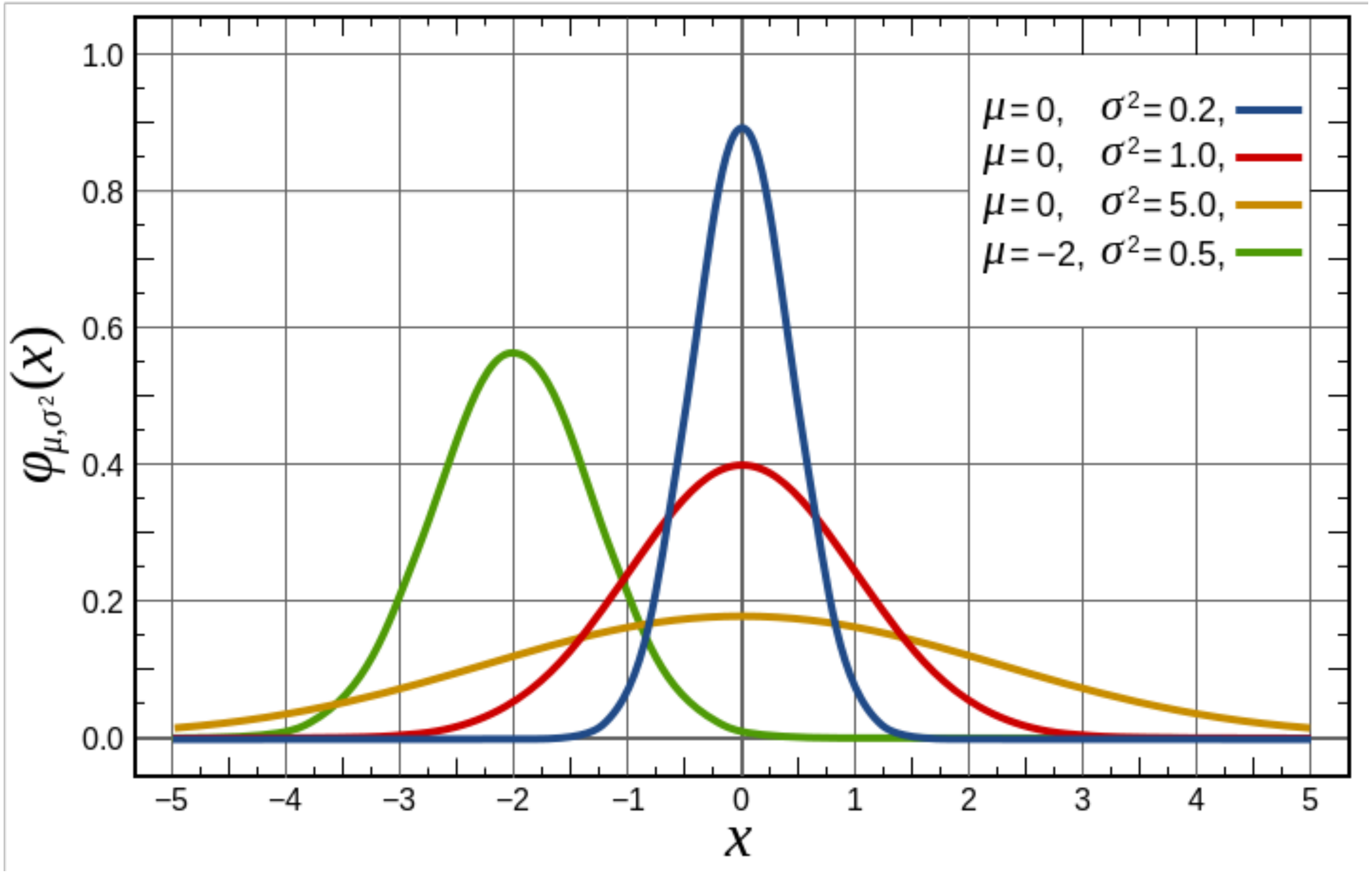
Definition

- A continuous random variable is said to be normally distributed with parameters mean μ and standard deviation σ if it has a probability density function given as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \text{ denoted as}$$

$$\mathcal{N}(\mu, \sigma^2)$$

- This pdf is symmetric about the mean *μ* and has the shape of the “bell curve”.



https://upload.wikimedia.org/wikipedia/commons/7/74/Normal_Distribution_PDF.svg

Definition

- If μ=0 and σ=1, it is called the **standard normal distribution**

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \text{ denoted as } \mathcal{N}(0,1)$$

- To verify that this is a valid pdf:

$$\left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)} dxdy = \int_0^{2\pi}\int_0^{\infty} e^{-r^2} r dr d\theta$$

Note the change from (x,y) to polar coordinates (r,θ).
x = r cos (θ)
y = r sin (θ)

$$= 2\pi \int_0^{\infty} e^{-r^2} r dr = \pi \int_0^{\infty} e^{-s} ds = \pi$$

$$\therefore \frac{1}{(1/\sqrt{2})\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2} dx = 1$$

This is a Gaussian pdf with mean 0 and standard deviation 1/sqrt(2). Thus we have verified that this particular Gaussian function is a valid pdf. You can verify that Gaussians with arbitrary mean and variance are valid pdfs by a change of variables.

Properties

- Mean:

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

$$=\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(y)e^{-(y)^2/(2)}dy$$

$$=0---why?$$

$$\therefore E[X]=\mu$$

Properties

- Variance:

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

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-(y)^2/(2)} dy$$

$$= \sigma^2 --why?$$

Properties

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

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

Proof on board. And in the book.

Properties

- Median = mean (why?)
- Because of symmetry of the pdf about the mean
- Mode = mean – can be checked by setting the first derivative of the pdf to 0 and solving, and checking the sign of the second derivative.
- CDF for a 0 mean Gaussian with variance 1 – is given by:

$$\Phi(x) = F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(x)^2/(2)} dx$$

Properties

- CDF – it is given by:

$$\Phi(x) = F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(x)^2/(2)} dx$$

- It is closely related to the error function erf(x) defined as:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-x^2} dx$$

- It follows that:

$$\Phi(x) = \frac{1}{2}\left[1 + erf\left(\frac{x}{\sqrt{2}}\right)\right]$$

Verify for yourself

Properties

- For a Gaussian with mean μ and standard deviation σ, it follows that:

$$\Phi\left(\frac{x-\mu}{\sigma}\right)=\frac{1}{2}\left[1+erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

- The probability that a Gaussian random variable has values from μ-*n*σ to μ+*n*σ is given by:

$$\Phi(\mu+n\sigma)-\Phi(\mu-n\sigma)=\frac{1}{2}erf\left(\frac{n}{\sqrt{2}}\right)-\frac{1}{2}erf\left(\frac{-n}{\sqrt{2}}\right)=erf\left(\frac{n}{\sqrt{2}}\right)$$

Properties

- The probability that a Gaussian random variable has values from μ-*n*σ to μ+*n*σ is given by:

$$\Phi(n) - \Phi(-n) = erf\left(\frac{n}{\sqrt{2}}\right)$$

n	Φ(*n*)-Φ(-*n*)
1	68.2%
2	95.4%
3	99.7%
4	99.99366%
5	99.9999%
6	99.9999998 %

Hence a Gaussian random variable lies within ±3σ from its mean with more than 99% probability

Properties

- MGF:

$$\Phi_X(t) = \exp\left(\mu t + \sigma^2 t^2 / 2\right)$$

Proof here.

A strange phenomenon

- Let's say you draw n = 2 values, called x_1 and x_2, from a [0,1] uniform random distribution and compute:

$$y_j = \sqrt{n}\left(\frac{\sum_{i=1}^{n} x_i}{n} - \mu\right)$$

Sampling index = i, $1 \leq i \leq n$
Iteration index = j, $1 \leq j \leq m$

(where μ is the true mean of the uniform random distribution)

- You repeat this process some m=5000 times (say), and then plot the histogram of the computed $\{y_j\}, 1 \leq j \leq m$, values.
- Now suppose you repeat the earlier two steps with larger and larger n.

A strange phenomenon

- Now suppose you repeat the earlier two steps with larger and larger *n*.
- It turns out that as *n* grows larger and larger, the histogram starts resembling a 0 mean Gaussian distribution with variance equal to that of the sampling distribution (i.e. the [0,1] distribution).
- Now if you repeat the experiment with samples drawn from any other distribution instead of [0,1] uniform random (i.e. you change the sampling distribution).
- The phenomenon still occurs, though the resemblance may start showing up at smaller or larger values of *n*.
- This leads us to a very interesting theorem called the **central limit theorem**.
- Demo code: http://www.cse.iitb.ac.in/~ajitvr/CS215_Fall2017/CLT/

Central limit theorem

- Consider $X_1, X_2, \ldots, X_n$ to be a sequence of independent and identically distributed (i.i.d.) random variables each with mean μ and variance $\sigma^2 < \infty$. Then as $n \to \infty$, the *distribution* (i.e. CDF) of the following quantity:

$$Y_n = \sqrt{n}\left(\frac{\sum_{i=1}^{n} x_i}{n} - \mu\right)$$

converges to that of $\mathcal{N}(0, \sigma^2)$. Or, we say Y_n converges in distribution to $\mathcal{N}(0, \sigma^2)$. This is called the **Lindeberg-Levy central limit theorem**.

Central limit theorem: some comments

- Note that the random variables $X_1, X_2, \ldots, X_n$ must be independent and identically distributed.

- Converge in distribution means the following:

$$\lim_{n\to\infty} P(Y_n \le z) = \Phi(z/\sigma)$$

- There is a version of the central limit theorem that requires only independence – and allows the random variables to belong to different distributions. This extension is called the **Lindeberg Central Limit theorem**, and is given on the next slide.

Lindeberg's Central limit theorem

- Consider $X_1, X_2, \ldots, X_n$ to be a sequence of independent random variables each with mean μ_i and variance $(\sigma_i)^2 < \infty$. Then as $n \to \infty$, the distribution of the following quantity:

$$Y_n = \left(\frac{\sum_{i=1}^{n} (x_i - \mu_i)}{\sum_{i=1}^{n} \sigma_i^2} \right)$$

$$\mathbf{1}_A : X \to \{0,1\}$$
$$\mathbf{1}_A(x) = 1 \text{ if } x \in A$$
$$= 0 \text{ otherwise}$$

Indicator function

converges to that of $\mathcal{N}(0, 1)$, provided for every $\varepsilon > 0$

$$\lim_{n\to\infty} \left(\frac{\sum_{i=1}^{n} E[(x_i - \mu_i)^2 . 1_{\{|x_i - \mu_i| > \varepsilon s_n\}}]}{s_n} \right) = 0, s_n = \sum_{i=1}^{n} \sigma_i^2$$

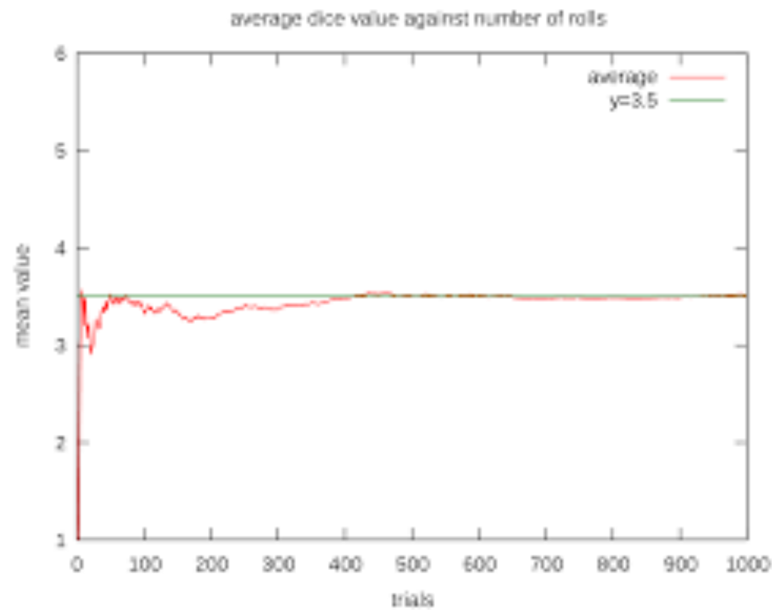
Lindeberg's Central limit theorem

- Informally speaking, the take home message from the previous slide is that the CLT is valid even if the random variables emerge from different distributions.
- This provides a major motivation for the widespread usage of the Gaussian distribution.
- The errors in experimental observations are often modelled as Gaussian – because these errors often stem from many different independent sources, and are modelled as being weighted combinations of errors from each such source.

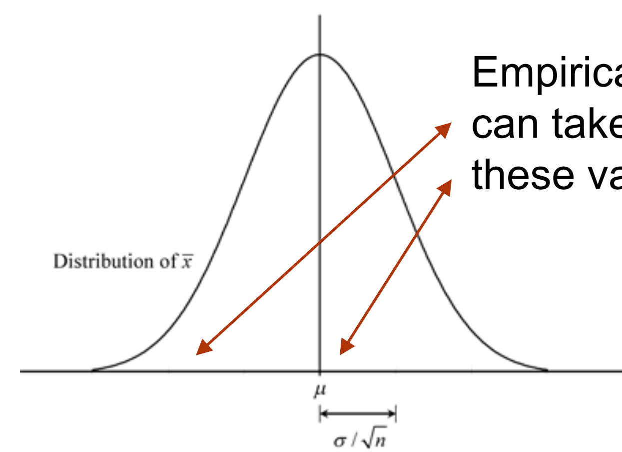
Central limit theorem versus law of large numbers

- The law of large numbers says that the empirical mean calculated from a large number of samples is **equal to** (or very close) to the *true mean* μ (of the distribution from which the samples were drawn).

- The central limit theorem says that the empirical mean calculated from a large number of samples is a **random variable** drawn from a Gaussian distribution with mean equal to the true mean μ (of the distribution from which the samples were drawn).



source



Empirical mean can take any of these values!

Central limit theorem versus law of large numbers

- Is this a contradiction?

Central limit theorem versus law of large numbers

- The answer is NO!
- Go and look back at the central limit theorem.

$$Y_n = \sqrt{n}\left(\frac{\sum_{i=1}^{n} x_i}{n} - \mu\right) \sim \mathcal{N}(0, \sigma^2)$$

$$\rightarrow \left(\frac{\sum_{i=1}^{n} x_i}{n} - \mu\right) \sim \mathcal{N}(0, \sigma^2 / n) \quad (--why?)$$

$$\rightarrow \left(\frac{\sum_{i=1}^{n} x_i}{n}\right) \sim \mathcal{N}(\mu, \sigma^2 / n)$$

This variance drops to 0 when *n* is very large! All the probability is now concentrated at the mean!

Proof of Lindberg-Levy CLT using MGFs

- Consider the *n* i.i.d. random variables $X_1, X_2, \ldots, X_n$ with mean μ and variance σ^2. Let their sum be S_n.
- Then we have to prove that:

$$\lim_{n\to\infty} P\left(\frac{S_n - \mu n}{\sigma\sqrt{n}} \le x\right) = \int_{-\infty}^{x} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy$$

- For that we will prove that the MGF of Z_n equals the MGF of the standard normal distribution (i.e. $\exp(t^2/2)$) where

$$Z_n = \left(\frac{S_n - \mu n}{\sigma\sqrt{n}}\right)$$

Proof of Lindberg-Levy CLT using MGFs

- By properties of the MGF, we have:

$$\phi_{S_n-\mu n}(t) = \left(\phi_X(t)\right)^n$$

$$\therefore \phi_{Z_n}(t) = \left(\phi_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n$$

$$\text{Recall}: \phi_Y(t) = e^{tb}\phi_X(at)$$

- We need to prove that:

$$\lim_{n\to\infty} n\log\left(\phi_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right) = t^2/2$$

Recall: for a Gaussian r.v. X with mean μ and std.dev. σ,

$$\Phi_X(t) = \exp\left(\mu t + \sigma^2 t^2/2\right)$$

Proof of Lindberg-Levy CLT using MGFs

- Labelling $x = 1/\sqrt{n}$, we have:

$$\lim_{x\to 0}\frac{\log\phi_X(tx/\sigma)}{x^2} = \lim_{x\to 0}\frac{\left(\dfrac{t\phi'_X(tx/\sigma)}{\sigma\phi_X(tx/\sigma)}\right)}{2x}$$

$$= \frac{t}{2\sigma}\lim_{x\to 0}\frac{\phi'_X(tx/\sigma)}{x\phi_X(tx/\sigma)}$$

$$= \frac{t}{2\sigma}\lim_{x\to 0}\frac{(t/\sigma)\phi''_X(tx/\sigma)}{\phi_X(tx/\sigma)+(t/\sigma)x\phi'_X(tx/\sigma)}$$

$$= \frac{t^2}{2\sigma^2}\frac{\phi''_X(0)}{\phi_X(0)} = \frac{t^2}{2\sigma^2}\frac{E(X^2)}{E(X^0)} = \frac{t^2\sigma^2}{2\sigma^2} = \frac{t^2}{2}$$

L'Hospital's rule

Recall:

$$\forall r\geq 0, \phi_X^{(r)}(0) = E(X^r)$$
$$\phi_X(0) = 1, \phi_X'(0) = 0$$

Application

- Your friend tells you that in 10000 successive independent unbiased coin tosses, he counted 5200 heads. Is (s)he serious or joking?

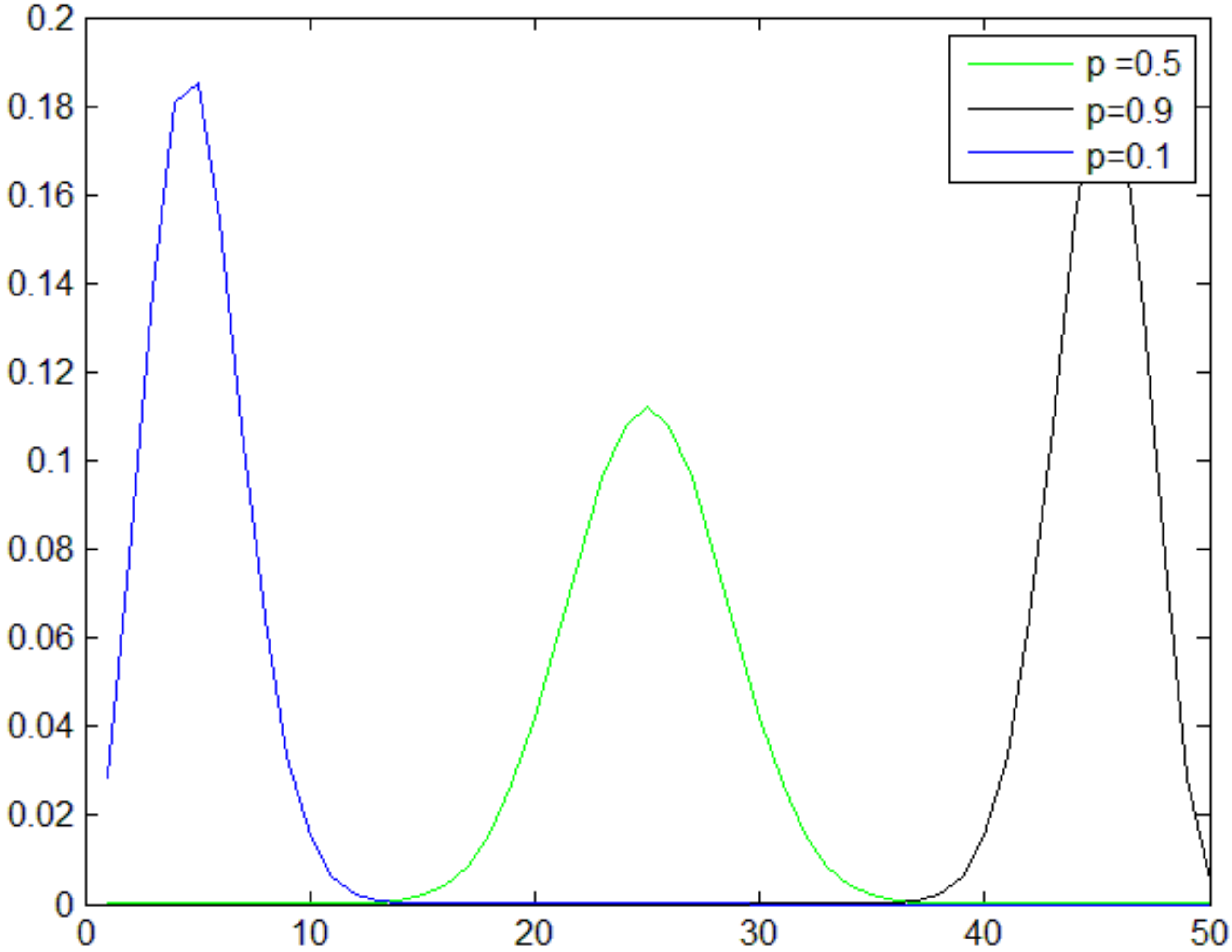
- **Answer:** Let X_1, X_2,…, X_n be the random variables indicating whether or not the coin toss was a success (a heads).

- These are i.i.d. random variables whose sum is a random variable with mean $n\mu$=10000(0.5) = 5000 and standard deviation $\sigma n^{1/2}$ = sqrt(0.5(1-0.5))sqrt(10000) = 50.

Application

- Your friend tells you that in 10000 successive independent unbiased coin tosses, he counted 5200 heads. Is (s)he serious or joking?
- **Answer:** The given number of heads is 5200 which is 4 standard deviations away from the mean.
- The chance of that occurring is of the order of 0.00001 (see the slide on error functions) since the total number of heads is a Gaussian random variable (as per central limit theorem).
- So your friend is (most likely) joking.
- Notice that this answer is much more principled than giving an answer purely based on some arbitrary threshold over $|X\text{-}5000|$.
- You will study much more of this when you do a topic called hypothesis testing.

Binomial distribution and Gaussian distribution



Binomial distribution and Gaussian distribution

- The binomial distribution begins to resemble a Gaussian distribution with an appropriate mean for large values of *n*.

- In fact this resemblance begins to show up for surprisingly small values of *n*.

- Recall that a binomial random variable is the number of successes of independent Bernoulli trials

$$X = \sum_{i=1}^{n} X_i, X_i = 1\,(\text{heads on } i^{\text{th}} \text{ trial}) \text{ else } 0$$

Binomial distribution and Gaussian distribution

- Each X_i has a mean of p and standard deviation of $p(1-p)$.

- Hence the following random variable is a standard normal random variable by CLT:

$$\frac{X - np}{\sqrt{np(1-p)}}$$

- Watch the animation here.

Binomial distribution and Gaussian distribution

- Another way of stating the afore-mentioned facts is that:

When $n \to \infty$, we have $\forall a, b, a \leq b$,

$$P\left(a \leq \frac{X - np}{\sqrt{np(1-p)}} \leq b\right) = \Phi(b) - \Phi(a)$$

where $X \sim Binomial(n, p)$

- This is called the **de Moivre-Laplace theorem** and is a special case of the CLT. But its proof was published almost 80 years before that of the CLT!

Distribution of the sample mean

- Consider independent and identically distributed random variables $X_1, X_2, \ldots, X_n$ with mean μ and standard deviation σ.

- We know that the sample mean (or empirical mean) is a ***random variable*** given by:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Note yet again: The true mean μ is NOT a random variable. The sample mean is, and its value converges to the true mean μ by the law of large numbers.

Distribution of the sample mean

- Now we have:

$$E(\overline{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \mu$$

$$Var(\overline{X}) = \frac{\sum_{i=1}^{n} Var(X_i)}{n^2} = \frac{\sigma^2}{n}$$

If $X_1, X_2, ..., X_n$ were normal random variables,
then it can be proved that $\overline{X}$ is also a normal random variable (how?).
Otherwise if $X_1, X_2, ..., X_n$ weren't normal random variables,
$\overline{X}$ would be only ***approximately*** normally distributed, as per
the central limit theorem.

Distribution of the sample variance

- The sample variance is given by:

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1} = \frac{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}{n-1}$$

- The sample standard deviation is *S*.

Distribution of the sample variance

- The expected value of the sample variance is derived as follows:

$$E((n-1)S^2) = E(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2)$$

$$= nE(X_1^2) - nE(\overline{X}^2)$$

$$E(W^2) = Var(W) + (E(W))^2$$

$$\therefore E((n-1)S^2) = nVar(X_1) + n(E(X_1))^2$$

$$- nVar(\overline{X}) - n(E(\overline{X}))^2$$

Distribution of the sample variance

- The expected value of the sample variance is derived as follows:

$$\therefore E((n-1)S^2) = nVar(X_1) + n(E(X_1))^2 - nVar(\overline{X}) - n(E(\overline{X}))^2$$

$$\therefore E((n-1)S^2) = n\sigma^2 + n\mu^2 - n(\sigma^2 / n) - n(\mu)^2 = (n-1)\sigma^2$$

$$\therefore E(S^2) = \sigma^2$$

Distribution of the sample variance

- The expected value of the sample variance is derived as follows:

$$\therefore E((n-1)S^2) = (n-1)\sigma^2$$

$$\therefore E(S^2) = \sigma^2$$

If the sample variance were instead defined as

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n} = \frac{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2}{n},$$

we would have :

$$E(S^2) = \frac{(n-1)\sigma^2}{n}$$

This is **undesirable** – as we would like to have the expected value of the sample variance to equal the true variance! Hence S^2 here above is multiplied by (n-1)/n to correct for this anomaly giving rise to our strange definition of sample variance. This multiplication by (n-1)/n is called **Bessel's correction.**

Distribution of the sample variance

- But the mean and the variance alone does not determine the distribution of a random variable.

- So what about the distribution of the sample variance?

- For that we need to study another distribution first – the **chi-squared distribution**.

Chi-square distribution

- If Z_1, Z_2, …., Z_n are *independent* standard normal random variables, then the following quantity

$$X = Z_1^2 + Z_2^2 + \ldots + Z_n^2$$

is said to have a chi-square distribution with *n* degrees of freedom and is denoted as follows

$$X \sim \chi_n^2$$

- The formula for this is as follows:

$$f_X(x) = \frac{x^{n/2-1} e^{-x/2}}{2^{n/2}\Gamma(n/2)}$$

$$\Gamma(y) = (y-1)!\,(y \text{ integer})$$

$$= \int_0^\infty x^{y-1} e^{-x} dx \,(y \text{ real})$$

Chi-square distribution

- To obtain the expression for the chi-square distribution when n = 1:

$$
\begin{aligned}
&Z_1 \sim N(0,1) \\
&X = Z_1^2 \\
&F_X(x) = P(Z_1^2 \le x) = P(-\sqrt{x} \le Z_1 \le \sqrt{x}) \\
&= F_{Z_1}(\sqrt{x}) - F_{Z_1}(-\sqrt{x}) \\
&f_X(x) = \frac{1}{2\sqrt{x}} f_{Z_1}(\sqrt{x}) + \frac{1}{2\sqrt{x}} f_{Z_1}(-\sqrt{x}) \\
&= \frac{1}{2\sqrt{x}} \frac{e^{-x/2}}{\sqrt{2\pi}} \times 2 = \frac{1}{\sqrt{x}} \frac{e^{-x/2}}{\sqrt{2\pi}}
\end{aligned}
$$

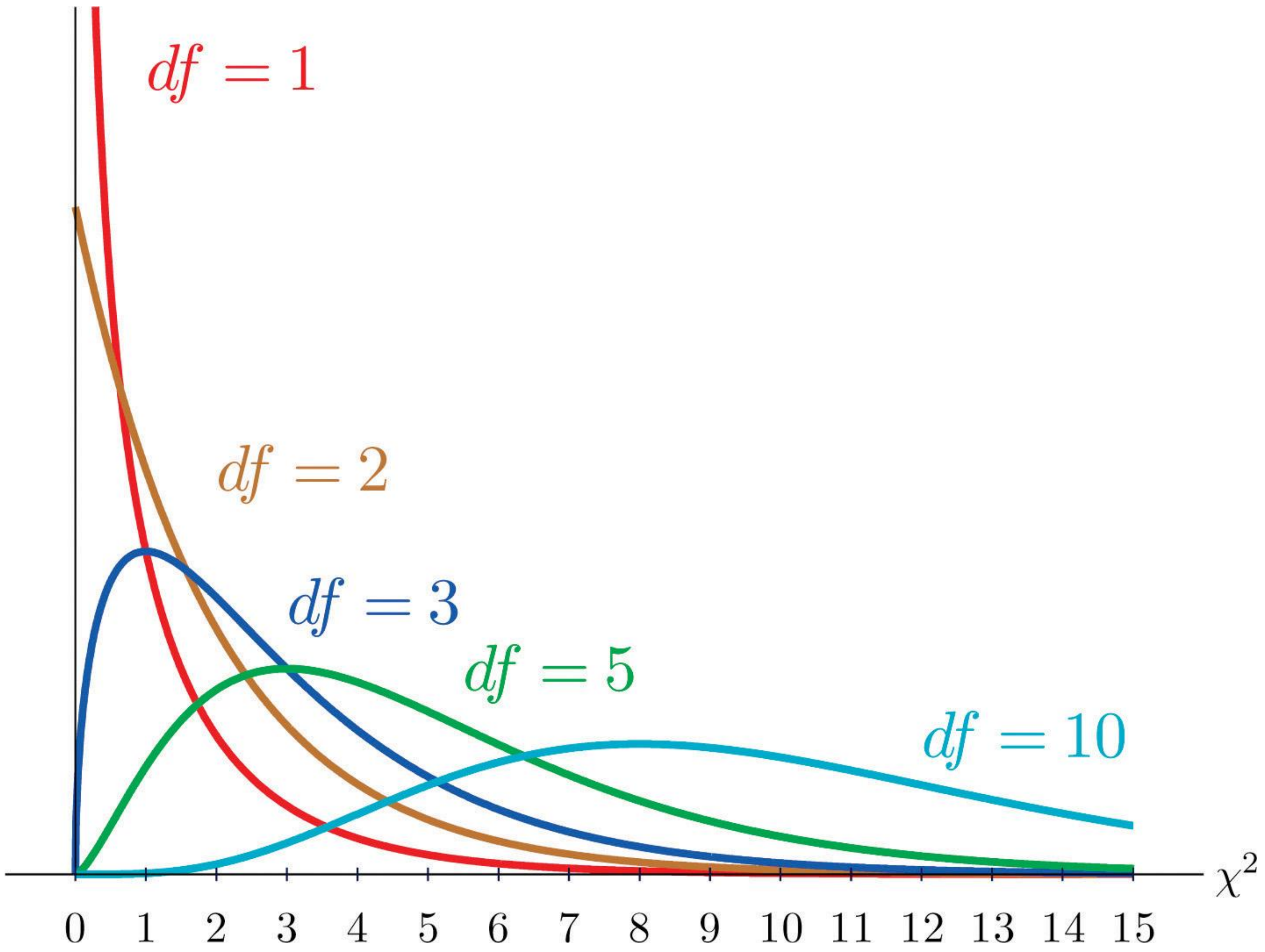
Chi-square distribution

- MGF of a chi-square distribution with *n* degrees of freedom:

$$\Phi_X(t) = (1-2t)^{-n/2}$$

Proof on the board. And here.

- Please note that the aforementioned MGF is defined only for $t < \frac{1}{2}$.



$df = 1$
$df = 2$
$df = 3$
$df = 5$
$df = 10$
χ^2
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Additive property

- If X_1 and X_2 are independent chi-square random variables with n_1 and n_2 degrees of freedom respectively, then $X_1 + X_2$ is also a chi-square random variable with $n_1 + n_2$ degrees of freedom. This is called the **additive property**.

- It is easy to prove this property by observing that $X_1 + X_2$ is basically the sum of $n_1 + n_2$ independent normal random variables.

Chi-square distribution

- Tables for the chi-square distribution are available for different number of degrees of freedom, and for different values of the independent variable.

Back to the distribution of the sample variance

$$(n-1)S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\overline{X} - \mu)^2$$

$$\therefore \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} + \frac{n(\overline{X} - \mu)^2}{\sigma^2}$$

$$\therefore \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} + \left(\frac{\sqrt{n}(\overline{X} - \mu)}{\sigma}\right)^2$$

The sum of squares of *n* standard normal random variables

The square of a standard normal random variable

Back to the distribution of the sample variance

$$\therefore \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} + \left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right)^2$$

The sum of squares of *n* standard normal random variables

The square of a standard normal random variable

It turns out that these two quantities are **independent** random variables. The proof of this requires multivariate statistics and transformation of random variables, and is deferred to a later point in the course. If you are curious, you can browse this link, but it's not on the exam for now.

Given this fact about independence, it then follows that the middle term is a chi-square distribution with *n*-1 degrees of freedom.

Uniform distribution

Uniform distribution

- A uniform random variable over the interval [a,b] has a pdf given by:

$$f_X(x) = 1/(b-a), \mathit{if}\ a \le x \le b$$
$$= 0 \text{ otherwise}$$

- Clearly, this is a valid pdf – it is non-negative and integrates to 1.
- It is easy to show that its mean and median are equal to $(b+a)/2$.

$$E(X) = \int_a^b \frac{xdx}{b-a} = \frac{x^2 \big|_a^b}{2(b-a)} = \frac{a+b}{2}$$

Uniform distribution

- Variance:

$$E(X^2) = \int_a^b \frac{x^2 dx}{b-a} = \frac{x^3 \big|_a^b}{3(b-a)} = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + a^2 + ab}{3}$$

$$Var(X) = E(X^2) - (E(X))^2 = \frac{(b-a)^2}{12}$$

- MGF:

$$E(e^{tX}) = \int_a^b \frac{e^{tx} dx}{b-a} = \frac{e^{tb} - e^{ta}}{t(b-a)}, t \neq 0$$

$$= 1, t = 0$$

Applications

- Uniform random variables, especially over the [0,1] interval are very important, in developing programs to draw samples from other distributions including the Gaussian, Poisson, and others.

- You will study more of this later on in the semester.

- For now, we will study two applications. How do you draw a sample from a distribution of the following form:

$$P(X = x_i) = p_i, 1 \leq i \leq n,$$

$$\sum_{i=1}^{n} p_i = 1$$

Applications

- For now, we will study two applications. How do you draw a sample from a discrete distribution with the following pmf:

$$P(X = x_i) = p_i, 1 \le i \le n, \sum_{i=1}^{n} p_i = 1$$

$$\begin{aligned}
&\text{Draw } u \sim Uniform(0,1) \\
&\text{If } u \le p_1 \rightarrow \text{sampled value is } x_1 \\
&\text{If } p_1 \le u \le p_1 + p_2 \rightarrow \text{sampled value is } x_2 \\
&. \\
&. \\
&\text{If } p_1 + p_2 + \ldots + p_{n-1} \le u \le p_1 + p_2 + \ldots + p_{n-1} + p_n \rightarrow \text{sampled value is } x_n
\end{aligned}$$

Applications

- Uniform random variables, especially over the [0,1] interval are very important, in developing programs to draw samples from other distributions including the Gaussian, Poisson, and others.

- You will study more of this later on in the semester.

- For now, we will study how to generate a random permutation of *n* elements. That is what the **"randperm"** function in MATLAB does, and you have used it at least once so far!

Application: generating a random subset

- In fact, we will do something more than randperm – we will develop theory to generate a random subset of size *k* from a set $A=\{a_1, a_2, \ldots, a_n\}$ of size *n*, assuming all the C(*n*,*k*) subsets are equally likely.

- Let us define the following for each element *j* ($1 \leq j \leq n$):

$$I_j = 1 \text{ if } a_j \in B_k, \text{ else } 0$$

Notation for chosen subset (of size *k*)

- Now we will sequentially pick each element of the subset randomly as follows:

Application: generating a random subset

- Notice that $P(I_1=1) = k/n$. (why? Because there $C(n,k)$ ways to pick k objects out of n. Let us say object 1 is one of them, then there are $C(n-1,k-1)$ ways to pick the other $k-1$ objects)

- If I_1 is 1, then $P(I_2=1) = (k-1)/(n-1)$. (why?)

- If I_1 is 0, then $P(I_2=1) = k/(n-1)$. (why?)

- Thus $P(I_2=1|I_1) = (k-I_1)/(n-1)$ (why?)

- Side question: what is $P(I_2=1)$?

Application: generating a random subset

- Continuing this way, one can show that:

$$P(I_j \mid I_1, I_2, \ldots, I_{j-1}) = \frac{k - \sum_{i=1}^{j-1} I_i}{n - (j-1)}, 2 \le j \le n$$

Application: generating a random subset

- This suggests the following procedure:

$$U_1 \sim Uniform(0,1)$$
$$I_1 = 1, \text{if } U_1 < k/n, \text{else } 0$$
$$U_2 \sim Uniform(0,1)$$
$$I_2 = 1, \text{if } U_2 < (k - I_1)/(n-1), \text{else } 0$$

.

.

When does this process stop? It stops at step #*j*

* If $I_1+I_2+\ldots+I_j = k$ and the random subset B_k contains those indices whose I-values are 1

OR

* If the number of unfilled entries in the random subset B_k = number of remaining elements in *A*. In this case B_k = all remaining elements in *A* with index greater than *i*=largest index in B_k. See figure 5.6 of the book.

$$U_j \sim Uniform(0,1)$$
$$I_j = 1, \text{if } U_j < (k - I_1 - I_2 - \ldots - I_{j-1})/(n - j + 1), \text{else } 0$$

Exponential Distribution

Motivation

- Consider a Poisson distribution with an average number of successes per unit time given by λ.
- So the number of successes in time u is λu.
- This is actually called a **Poisson process**.
- Now consider the time taken (T) for the first success – this is called as the **waiting time**.

Motivation

- Let X ~ Poisson (λu) for time interval $(0, u)$.

- T is a random variable whose distribution we are going to seek to model here. Then,

$$P(T \le u) = 1 - P(T > u) = 1 - P(X = 0)$$

The probability that the first success occurred after time t = probability that there was no success in the time interval $(0, t)$, i.e. $X = 0$ in that interval

$$P(T \le u) = 1 - P(T > u) = 1 - P(X = 0)$$

$$= 1 - \frac{e^{-\lambda u}(\lambda u)^0}{0!} = 1 - e^{-\lambda u}$$

Motivation

$$F_T(u) = 1 - e^{-\lambda u}$$

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

$$(0 \text{ elsewhere})$$

Such a random variable *T* is called an exponential random variable. It models the waiting time for a Poisson process. It has a parameter λ.

Properties: MGF

- MGF defined for $t < \lambda$:

$$\Phi_T(t) = E(e^{tT})$$

$$= \int_0^\infty e^{tu} \lambda e^{-\lambda u} du = \lambda \int_0^\infty e^{-(\lambda - t)u} du = \frac{\lambda}{\lambda - t}$$

Properties: Mean and Variance

- Mean:

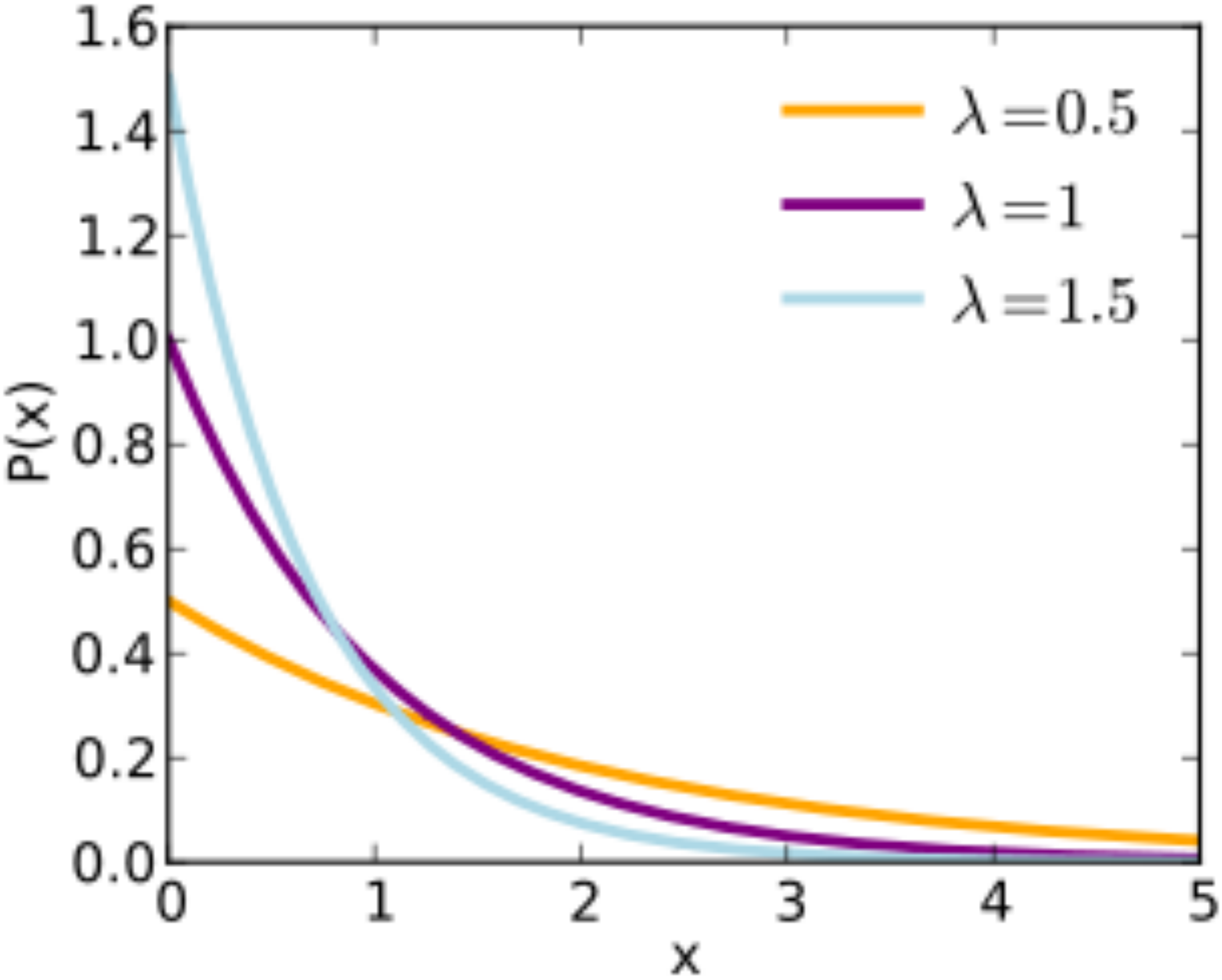
$$E(T) = \int_0^\infty t\lambda e^{-\lambda t} dt = \lambda \int_0^\infty t e^{-\lambda t} dt = \frac{1}{\lambda}$$

This is intuitive – a Poisson process with a large average rate should definitely lead to a lower expected waiting time.

- Variance:

$$Var(T) = E(T^2) - (E(T))^2 = \frac{1}{\lambda^2}$$

$$E(T^2) = \int_0^\infty t^2 \lambda e^{-\lambda t} dt = \frac{2}{\lambda^2}$$


1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
P(x)
0
1
2
3
4
5
x
λ=0.5
λ=1
λ=1.5

Properties: Mode and Median

- Mode: always at 0

- Median

$$\int_0^u \lambda e^{-\lambda t} dt = \frac{1}{2}$$

$$u = \frac{\ln 2}{\lambda}$$

Properties: "Memorylessness"

- A non-negative random variable *T* is said to be **memoryless** if:

$$\forall s, u \geq 0, P(T > s + u \mid T > u) = P(T > s)$$

- Meaning: This gives the probability that given a waiting time of success of more than *u*, the waiting time will exceed *s*+*u*, i.e. one would have to wait for *s* more time units for success.

- Another formula (equivalent to the earlier one)

$$\frac{P(T > s + u, T > u)}{P(T > u)} = P(T > s)$$

Properties: "Memorylessness"

- You can easily verify that this holds for the exponential distribution.

$$P(T > u) = e^{-\lambda u}$$

$$P(T > s) = e^{-\lambda s}$$

$$P(T > s + u) = e^{-\lambda(s+u)}$$

Example

- Suppose that the number of miles a car can run before its battery fails is exponentially distributed with an average of α. What is the probability that the car won't fail on a trip of *k* miles given that it has already run for *l* miles?

- Solution: For exponential distribution we know that

$$P(T > k + l \mid T > l) = \frac{P(T > k + l, T > l)}{P(T > l)}$$

$$= P(T > k) = e^{-k\lambda} = e^{-k/\alpha}$$

Example

- Suppose that the number of miles a car can run before its battery fails is exponentially distributed with an average of α. What is the probability that the car won't fail on a trip of k miles given that it has already run for *l* miles?

- Solution: If the distribution were not exponential, then we have

$$P(T > k + l \mid T > l) = \frac{P(T > k + l, T > l)}{P(T > l)} = \frac{1 - F_T(k + l)}{1 - F_T(l)}$$

Property: Minimum

- Consider independent exponentially distributed random variables $X_1, X_2, \ldots, X_n$ with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then $\min(X_1, X_2, \ldots, X_n)$ is exponentially distributed.

$$P(\min(X_1, X_2, \ldots, X_n) > x) = P(X_1 > x, X_2 > x, \ldots, X_n > x)$$

$$= \prod_{i=1}^{n} P(X_i > x) \text{ due to independence}$$

$$= \prod_{i=1}^{n} e^{-\lambda_i x}$$

$$= e^{-\sum_{i=1}^{n} \lambda_i x}$$