CS70: Lecture 31.

Gaussian RVs and CLT

1. Review: Continuous Probability: Geometric and Exponential
2. Normal Distribution
3. Central Limit Theorem
4. Examples

Continuous Probability

1. pdf: $Pr[X \in (x, x+\delta]] = f_X(x)\delta$.
2. CDF: $Pr[X \leq x] = F_X(x) = \int_{-\infty}^{x} f_X(y)dy$.
3. $U[a,b]$, $Expo(\lambda)$, target.
4. Expectation: $E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$.
5. Expectation of function: $E[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x)dx$.
6. Variance: $var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$.
7. Variance of Sum of Independent RVs: If X_n are pairwise independent, $var[X_1 + \cdots + X_n] = var[X_1] + \cdots + var[X_n]$

Geometric and Exponential: Relationship - Recap

The geometric and exponential distributions are similar. They are both memoryless.

Consider flipping a coin every $1/N$ second with $Pr[H] = p/N$, where $N \gg 1$.

Let X be the time until the first H.

Fact: $X \approx Expo(p)$.

Analysis: Note that

$$\begin{aligned} Pr[X > t] &\approx Pr[\text{first } Nt \text{ flips are tails}] \\ &= (1 - \frac{p}{N})^{Nt} \approx \exp\{-pt\}. \end{aligned}$$

Indeed, $(1 - \frac{a}{N})^N \approx \exp\{-a\}$.

Minimum of Independent *Expo* Random Variables

Minimum of Independent ***Expo*****.** Let $X = Expo(\lambda)$ and $Y = Expo(\mu)$ be independent RVs.

Recall that $Pr[X > u] = e^{-\lambda u}$. Then

$$\begin{aligned} Pr[\min\{X,Y\} > u] &= Pr[X > u, Y > u] = Pr[X > u]Pr[Y > u] \\ &= e^{-\lambda u} \times e^{-\mu u} = e^{-(\lambda+\mu)u}. \end{aligned}$$

This shows that $\min\{X,Y\} = Expo(\lambda+\mu)$.

Thus, the minimum of two independent exponentially distributed RVs is exponentially distributed.

Maximum of Two Exponentials

Let $X = Expo(\lambda)$ and $Y = Expo(\mu)$ be independent. Define $Z = \max\{X, Y\}$.

Calculate $E[Z]$.

We compute f_Z, then integrate.

One has

$$\begin{aligned} Pr[Z < z] &= Pr[X < z, Y < z] = Pr[X < z]Pr[Y < z] \\ &= (1 - e^{-\lambda z})(1 - e^{-\mu z}) = 1 - e^{-\lambda z} - e^{-\mu z} + e^{-(\lambda+\mu)z} \end{aligned}$$

Thus,

$$f_Z(z) = \lambda e^{-\lambda z} + \mu e^{-\mu z} - (\lambda + \mu)e^{-(\lambda+\mu)z}, \forall z > 0.$$

Hence,

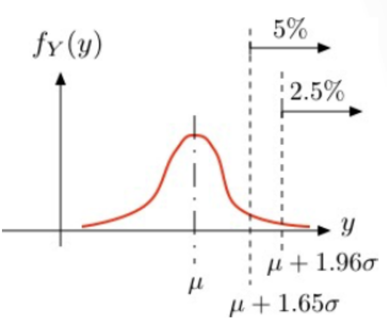
$$E[Z] = \int_0^\infty z f_Z(z) dz = \frac{1}{\lambda} + \frac{1}{\mu} - \frac{1}{\lambda + \mu}.$$

Normal (Gaussian) Distribution.

For any μ and σ, a **normal** (aka **Gaussian**) random variable Y, which we write as $Y = \mathcal{N}(\mu, \sigma^2)$, has pdf

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2}.$$

Standard normal has $\mu = 0$ and $\sigma = 1$.



Note: $Pr[|Y - \mu| > 1.65\sigma] = 10\%$; $Pr[|Y - \mu| > 2\sigma] = 5\%$.

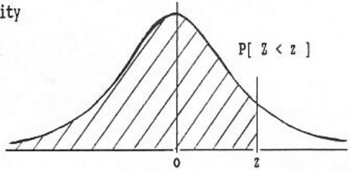
Table

STANDARD STATISTICAL TABLES

1. Areas under the Normal Distribution

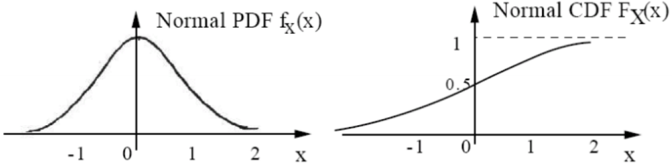
The table gives the cumulative probability up to the standardised normal value z i.e.

$$P[\, Z < z \,] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}Z^2)\, dZ$$



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	0.5159	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	0.7854
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	0.8804	0.8830

Standard Normal Variable



X is a standard normal variable if

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}, \quad x \in \Re$$

Since $f(x) = f(-x)$ and $\int_0^\infty x f_X(x) dx$ is finite, $E[X] = 0$.

Standard Normal Variable

$$var(X) = E[X^2] = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x^2 e^{\frac{-x^2}{2}} dx$$

Integration by parts, $du = e^{\frac{-x^2}{2}}$, $v = x^2$

$$\begin{aligned} var(X) &= \frac{1}{\sqrt{2\pi}} (-xe^{\frac{-x^2}{2}})\Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} dx \\ &= 0 + 1 \end{aligned}$$

Bell Curve with zero mean and unit variance.

Standard Normal Variable (You're not responsible for this!)

We need to show that

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

Use a trick: Let the value of the integral be A. Then

$$A^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy$$

Now use polar co-ordinates.

$$A^2 = \frac{1}{2\pi} \int_{0}^{\infty} \int_{0}^{2\pi} e^{-\frac{r^2}{2}} rd\theta dr$$

$$= \frac{1}{2\pi} \int_{0}^{\infty} 2\pi e^{-\frac{r^2}{2}} rdr$$

Substituting:

$$A^2 = -e^{\frac{r^2}{2}}]_0^{\infty} = 1$$

Scaling and Shifting

Theorem Let $X = \mathcal{N}(0,1)$ and $Y = \mu + \sigma X$. Then

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

Proof: $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}$. Now,

$$\begin{aligned} f_Y(y) &= \frac{1}{\sigma} f_X(\frac{y-\mu}{\sigma}) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(y-\mu)^2}{2\sigma^2}\}. \quad \square \end{aligned}$$

Gaussian RV: Computing Probabilities

Example: Let the temperature of a city in the winter be X in Celsius.
$X \sim N(2, 16)$. Find $P(-2 < X < 6)$.
Want to avoid computing the hairy integrals involved...Closed form of Normal CDF not available.
Procedure:

1. Convert to a standard normal: $\hat{X} = \frac{X-2}{4}$
2. Express in terms of probability problem of $\hat{X}$:

$$P(-2 < X < 6) = P(\frac{-2-2}{4} < \hat{X} < \frac{6-2}{4}) = P(-1 < \hat{X} < 1)$$

3. Look it up in a table of computed values of $F_{\hat{X}}(\hat{x})$
   We want to find $P(\hat{X} < 1) - P(\hat{X} < -1) = \phi(1) - \phi(-1)$. Tables typically only list positive values, so $\phi(-1) = 1 - \phi(1)$ by symmetry.
4. Answer: $2\phi(1) - 1 =$

Note: $2\phi(k) - 1$ tells you how likely it is that the outcome is within k standard deviations. $k = 2 : 0.9544$;k=3: 0.9974

Crown Jewel of Normal Distribution

Central Limit Theorem

For any set of independent identically distributed (i.i.d.) random variables X_i, define $A_n = \frac{1}{n}\sum X_i$ to be the "running average" as a function of n.

Suppose the X_i's have expectation $\mu = E(X_i)$ and variance σ^2.

Then the Expectation of A_n is μ, and its variance is σ^2/n.

Interesting question: What happens to the **distribution** of A_n as n gets large?

Note: We are asking this for **any arbitrary original distribution X_i!**

Central Limit Theorem

Central Limit Theorem

Let $X_1, X_2, \ldots$ be i.i.d. with $E[X_1] = \mu$ and $var(X_1) = \sigma^2$. Define

$$S_n := \frac{A_n - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Then,

$$S_n \to \mathcal{N}(0,1), \text{as } n \to \infty.$$

That is,

$$Pr[S_n \leq \alpha] \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-x^2/2} dx.$$

Proof: See EE126.

Note:

$$E(S_n) = \frac{1}{\sigma/\sqrt{n}}(E(A_n) - \mu) = 0$$

$$Var(S_n) = \frac{1}{\sigma^2/n} Var(A_n) = 1.$$

Implications of CLT

1. The Distribution of S_n wipes out all the information in the original information except for μ and σ^2.
2. If there are large number of small and independent factors, the aggregate of these factors will be normally distributed. E.g. Noise.
3. The Gaussian Distribution is very important – many problems involve sums of iid random variables and the only thing one needs to know is the mean and variance.

Example: Guessing numbers

Alice picks 100 numbers uniformly from $[0, 1]$. Bob must guess the sum within 2. What is the probability he is right if he guesses 55?
If X_i is Alice's i^{th} guess, it is difficult to calculate the CDF of $S_{100} = \sum_i X_i$.
$E[S_n] = 50, \mathrm{var}(S_n) = 100/12$.
So by the CLT, $\frac{S_{100}-50}{\sqrt{100/12}}$ is approx $N(0, 1)$. Thus

$$P(53 \leq S_{100} \leq 57) \approx \phi(\frac{7}{2.887}) - \phi(\frac{3}{2.887}) = \phi(2.4247) - \phi(1.039)$$

$$= .9925 - .9251 = 0.0649$$

Example: Counting Customers

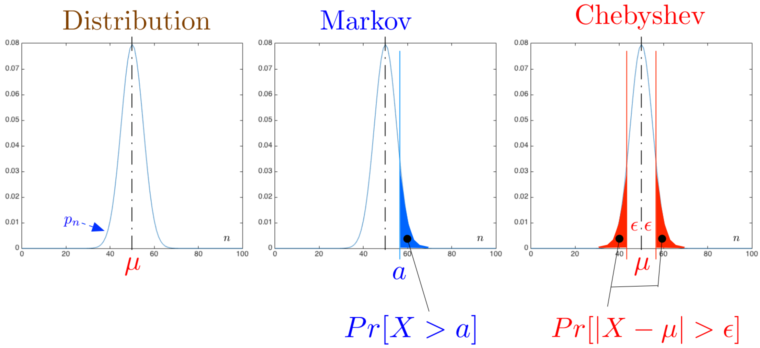
A store has one cashier who can serve one customer at a time.
The time taken to serve a customer $X \sim$ uniform on $[1, 5]$ minutes.
What is the probability that at least 200 customers can be served in 8 hours?
We want $P(\sum_{i=1}^{200} X_i \leq 480)$.
$E[S_{200}] = 600$, $var(S_{200}) = 200\ 16/12 = 800/3$.

$$P(200 \leq \sum_{i=1}^{200} X_i \leq 480) \approx \phi(\frac{480-600}{20\sqrt{2/3}}) - \phi(\frac{-400}{20\sqrt{2/3}})$$

$$= \phi(-6\sqrt{1.5}) - \phi(-20\sqrt{1.5}) = \phi(24.4949) - \phi(7.348).$$

$$P(\sum_{i=1}^{200} X_i \leq 480) \approx 0$$

Inequalities: A Preview



Summary

Gaussian and CLT

1. Gaussian: $\mathcal{N}(\mu, \sigma^2) : f_X(x) = ...$ "bell curve"
2. CLT: X_n i.i.d. $\implies \frac{A_n - \mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1)$