Math 170E: Winter 2023

Lecture 17, Wed 22nd Feb

The normal distribution and bivariate distributions of the discrete type

Example 11: The Normal distribution

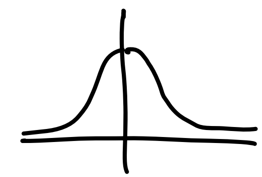
$\in \mathbb{R}$

- We say a continuous random variable X is normally distributed with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ if it has PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for } \; x \in \mathbb{R}$$

e^{-x^2}

Gaussian.



- We write $X \sim \mathcal{N}(\mu, \sigma^2)$

- If $\mu = 0$ and $\sigma^2 = 1$, we say that X is a standard normal random variable; i.e. $X \sim \mathcal{N}(0, 1)$, where

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

- **Last time:** We showed that

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

Proposition 3.17: If $X \sim \mathcal{N}(\mu, \sigma^2)$, then it has MGF

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \quad \text{for any} \quad t \in \mathbb{R}$$

$X \sim \mathcal{N}(0,1)$,
$M_X(t) = e^{\frac{1}{2}t^2}$.

Proof:

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx$$

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

$$tx - \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2} \underset{\text{complete the square}}{=} \cdots\cdots = -\frac{1}{2\sigma^2}\left(x - (\mu + \sigma^2 t)\right)^2 + \mu t + \frac{1}{2}\sigma^2 t^2.$$

$$= \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left(x - (\mu+\sigma^2 t)\right)^2\right) dx}_{=1} \cdot e^{\mu t + \frac{1}{2}\sigma^2 t^2}.$$

$$= e^{\mu t + \frac{1}{2}\sigma^2 t^2} \quad \text{for any } t \in \mathbb{R}.$$

Proposition 3.18: If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\left.\begin{aligned}\mathbb{E}[X] &= \mu \\ \text{var}(X) &= \sigma^2\end{aligned}\right\}$$

Proof: $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \rightarrow \log M_X(t) = \mu t + \frac{1}{2}\sigma^2 t^2$

$$\mathbb{E}[X] = \frac{d}{dt}\log M_X(t)\Big|_{t=0} = \mu + \sigma^2 t\Big|_{t=0} = \mu$$

$$\text{Var}(X) = \frac{d^2}{dt^2}\log M_X(t)\Big|_{t=0} = \sigma^2\Big|_{t=0} = \sigma^2 \quad //$$

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx =$$

Example 12:

- Let $X \sim \mathcal{N}(0,1)$ be a standard normal random variable
- What is $\mathbb{P}(X \leq 1.44)$?

$$\mathbb{P}(X \leq 1.44) = \int_{-\infty}^{1.44} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx.$$

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

if $X \sim N(0,1)$

In general, $\Phi(x)$ cannot be computed in terms of "elementary functions".

- We define the function

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

- That is, $\mathbb{P}(X \leq x) = \Phi(x)$, for $X \sim \mathcal{N}(0,1)$
- To find $\Phi(x)$ for $x > 0$, we use a table of standard normal CDF values:

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

$F_X(x) = \mathbb{P}(X \leq x)$
$= \Phi(x).$

$\mathbb{P}(X \leq 1.44)$
$= \Phi(1.44)$
$= \Phi(1.40 + 0.04)$
$= 0.9251.$

- Standard normal tables only have values for $x \geq 0$
- To obtain values for $x < 0$, we use the following result:

$$\mathbb{P}(X \leq -1.44) = \Phi(-1.44) = 1 - \underbrace{\Phi(1.44)}$$

Proposition 3.19: If $x \geq 0$, then

$$\Phi(-x) = 1 - \Phi(x)$$

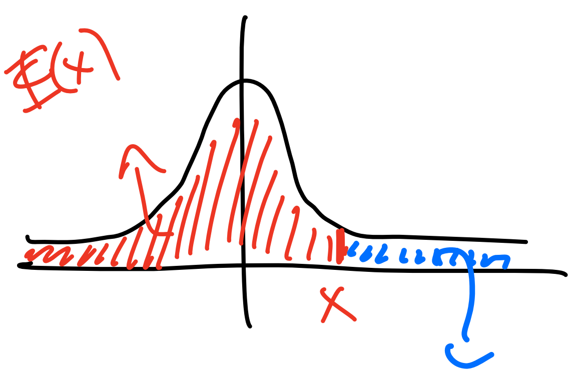
Proof:

$$\Phi(-x) = \int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \overset{u=-y}{=} \int_{+\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} (-du)$$

$$= \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \quad x \geq 0.$$

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

$$= 1 - \Phi(x).$$



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

- Standard normal tables only have values for standard normal distributions
- How do we get values for general normal distributions?

Proposition 3.19: If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

In particular, $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\mathbb{P}(X \leq x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Proof:

① $F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq z\right).$

$$= \mathbb{P}(X \leq \mu + \sigma z).$$

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

$F(\mu+\sigma z) \to \frac{d}{dz}(\mu+\sigma z) \times F'(\mu+\sigma z)$

② $f_Z(z) = F_Z'(z) = \frac{d}{dz}\left(\int_{-\infty}^{\mu+\sigma z} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right)$

$= \frac{d}{dz}(\mu+\sigma z) \cdot \frac{d}{dx}\left(\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \right) \Bigg|_{x=\mu+\sigma z}$

$= \cancel{\sigma} \cdot \frac{1}{\sqrt{2\pi\cancel{\sigma^2}}} e^{-\frac{(\cancel{\mu}+\cancel{\sigma} z-\cancel{\mu})^2}{2\cancel{\sigma^2}}}$

$= \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$

$\Rightarrow Z \sim N(0,1).$ $\quad Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

$\hookrightarrow X = \mu + \sigma Z.$

$\mathbb{P}(X \le x) = \mathbb{P}(\mu + \sigma Z \le x)$

$= \mathbb{P}\left(Z \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

$\hookrightarrow$ since $Z \sim N(0,1)$.

Example 12:

$\mu=4,\ \sigma^2=16 \Rightarrow \sigma=4$

- Let $X \sim \mathcal{N}(4, 16)$

- What is $\mathbb{P}(4 < X \leq 8)$?

$Z = \frac{X-4}{4} \sim N(0,1)$.

$\hookrightarrow X = 4 + 4Z$.

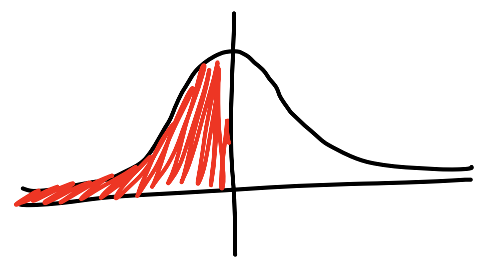
$$\mathbb{P}(4 < X \leq 8) = \mathbb{P}(4 < 4+4Z \leq 8).$$

$$= \mathbb{P}(1 < 1+Z \leq 2).$$

$$= \mathbb{P}(0 < Z \leq 1).$$

$$= \mathbb{P}(Z \leq 1) - \mathbb{P}(Z \leq 0).$$

$$= \Phi(1) - \Phi(0).$$



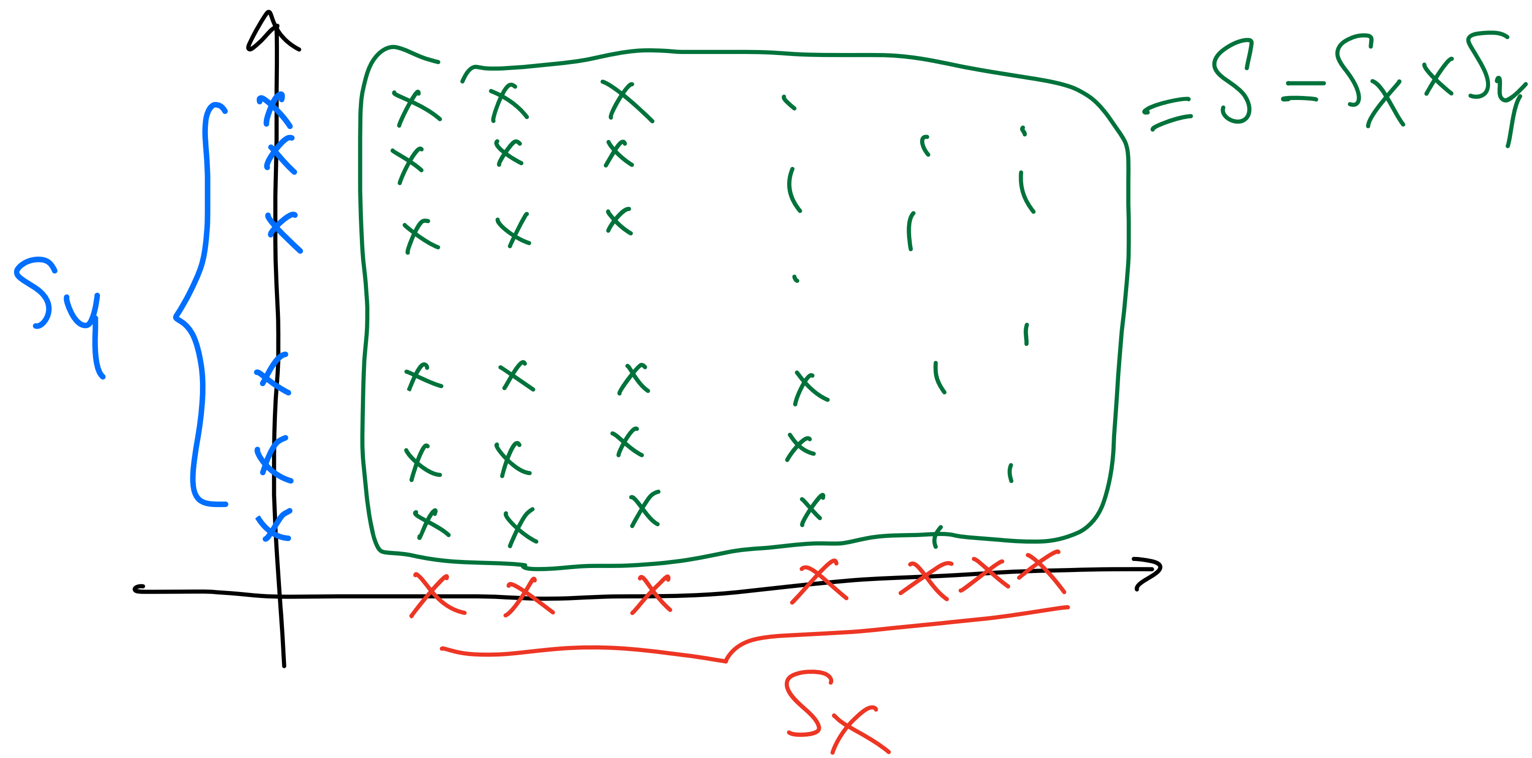
$$= 0.8413 - \Phi(0).$$

$$= 0.8413 - 0.5 = 0.3413 \ldots$$

Discrete bivariate distributions

pair (X, Y)

Definition 4.1: Let X, Y be a pair of discrete random variables taking values in sets $S_X, S_Y \subset \mathbb{R}$, respectively and let $S = S_X \times S_Y := \{(x,y) \in \mathbb{R}^2 : x \in S_X, y \in S_Y\}$.



- We define the joint probability mass function of X, Y to be the function $p_{X,Y} : S \to [0,1]$ by

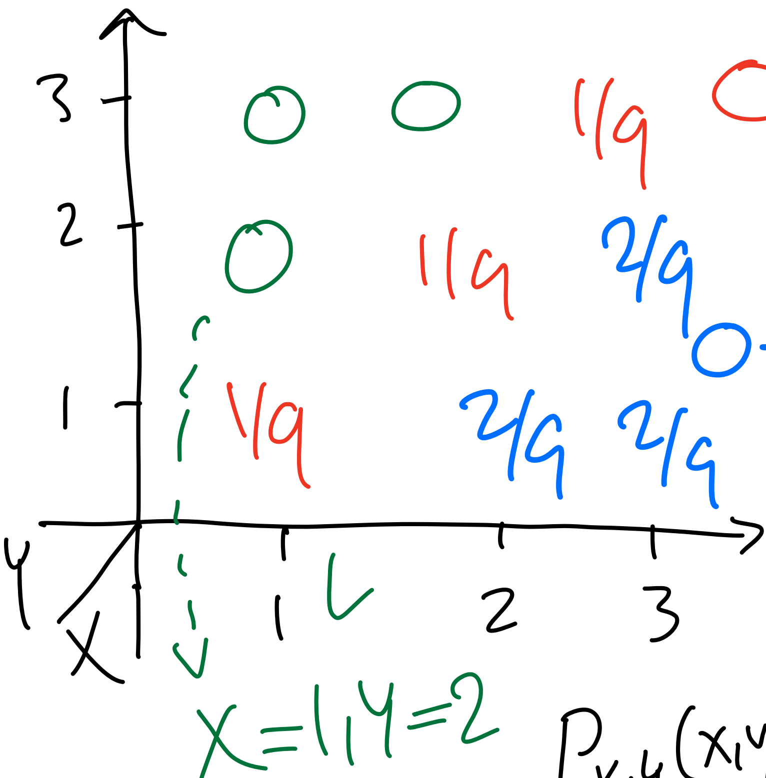
$\hookrightarrow$ Joint PMF.

$$p_{X,Y}(x,y) = \mathbb{P}(X = x,\ Y = y)$$
$$= \mathbb{P}((X,Y) = (x,y)).$$
$$= \mathbb{P}(\{X = x\} \cap \{Y = y\}).$$

Example 1: $\Omega = \{(1,1),(1,2),(1,3),(2,3), - - - \} \hookrightarrow |\Omega| = 3\times 3 = 9$

- You choose two numbers at random from the set $\{1, 2, 3\}$
- Let X be the larger and Y be the smaller of these two numbers
- What is the joint PMF of X, Y? $\quad (X,Y) \in S = \{1,2,3\}\times\{1,2,3\}$

$$P_{X,Y}(x,y) = \mathbb{P}(X=x, Y=y).$$



$$0 \to P_{X,Y}(1,1) = \mathbb{P}(X=1, Y=1) = \mathbb{P}(\{(1,1)\}) = \frac{1}{9}$$

$$0 \to P_{X,Y}(2,1) = \mathbb{P}(X=2, Y=1) = \mathbb{P}(\{(2,1),(1,2)\}) = 2/9$$

$$P_{X,Y}(x,y) = \begin{cases} 1/9 & \text{if } 1 \le x = y \le 3 \\ 2/9 & \text{if } 1 \le y < x \le 3. \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 4.2: Let X, Y be a pair of discrete random variables taking values in sets $S_X, S_Y \subset \mathbb{R}$, respectively and let $S = S_X \times S_Y$.

If X, Y have joint PMF $p_{X,Y}(x, y)$ and $A \subseteq \mathbb{R}^2$, then

$$\mathbb{P}((X, Y) \in A) = \sum_{(x,y)\in A\cap S} p_{X,Y}(x, y)$$

Proof: Single r.v.: $\mathbb{P}(X \in A) = \sum_{x \in A \cap S_X} p_X(x)$.

$$\mathbb{P}((X,Y) \in A) = \mathbb{P}((X,Y) \in A \cap S).$$

$$= \mathbb{P}\Big(\bigcup_{(x,y)\in A\cap S} \{X=x, Y=y\} \Big).$$

countable additivity

$$= \sum_{(x,y)\in A\cap S} \mathbb{P}(\underbrace{X=x, Y=y}_{p_{X,Y}(x,y)}).$$

$\hookrightarrow$ Normalisation condition: Put $A = \mathbb{R}^2$,

$$1 = \mathbb{P}\left((X,Y) \in \mathbb{R}^2\right) = \sum_{(x,y) \in \underbrace{\mathbb{R}^2 \cap S}_{= S}} p_{X,Y}(x,y) .$$

$$\hookrightarrow \quad \boxed{1 = \sum_{(x,y) \in S} p_{X,Y}(x,y)}$$