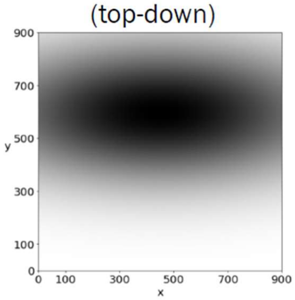
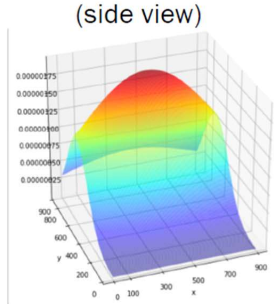
Back to darts

(top-down)



(side view)

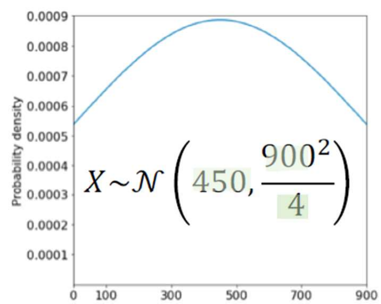


Darts were thrown according to a bivariate normal distribution:

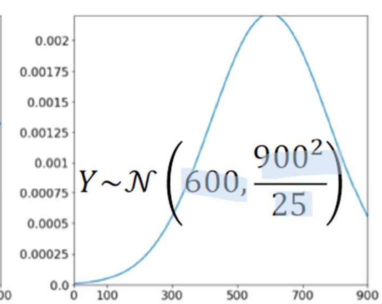
$$(X, Y) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \qquad \boldsymbol{\mu} = (450, 600) \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 900^2/4 & 0 \\ 0 & 900^2/25 \end{bmatrix}$$

says that $Cov(X_1, X_2) = 0$

Marginal PDFs:



pixel x



pixel y

A diagonal covariance matrix

Let $\boldsymbol{X} = (X_1, X_2)$ follow a bivariate normal distribution $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = (\mu_1, \mu_2), \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

still says $Cov(X_1, X_2) = 0$

Are X_1 and X_2 independent?

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\underbrace{\rho^2}_{0})}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \underbrace{\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}}_{\text{goes to zero}} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}$$

$$= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}$$

(Note covariance: $\rho\sigma_1\sigma_2 = 0$)

$$= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x_1-\mu_1)^2/2\sigma_1^2} \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(x_2-\mu_2)^2/2\sigma_2^2}$$ ✅

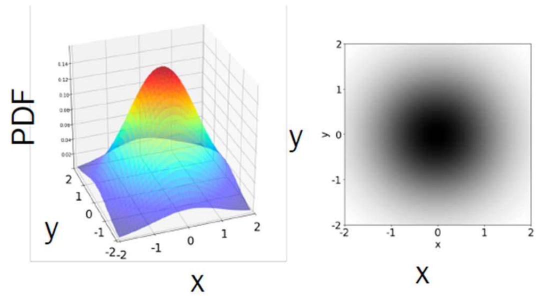
X_1 and X_2 are **independent** with marginal distributions $X_1 \sim \mathcal{N}(\mu_1\ \sigma_1^2), X_2 \sim \mathcal{N}(\mu_2\ \sigma_2^2)$

(X,Y) Matching (all have $\boldsymbol{\mu} = (0,0)$)

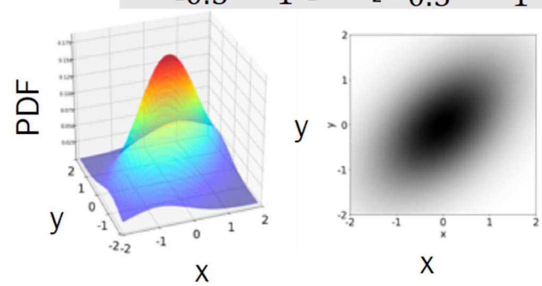
A. $\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$ B. $\begin{bmatrix}1 & 0\\0 & 2\end{bmatrix}$

C. $\begin{bmatrix}1 & 0.5\\0.5 & 1\end{bmatrix}$ D. $\begin{bmatrix}1 & -0.5\\-0.5 & 1\end{bmatrix}$

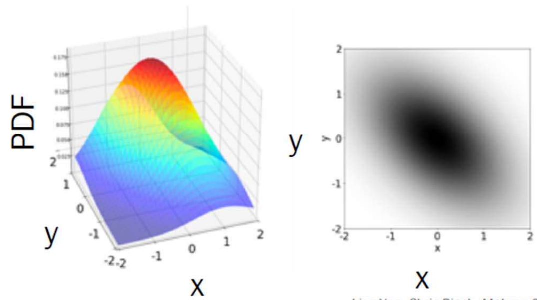
1.



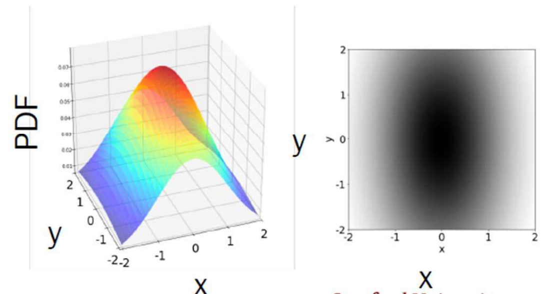
2.



3.



4.



(X,Y) Matching (all have $\boldsymbol{\mu} = (0,0)$)

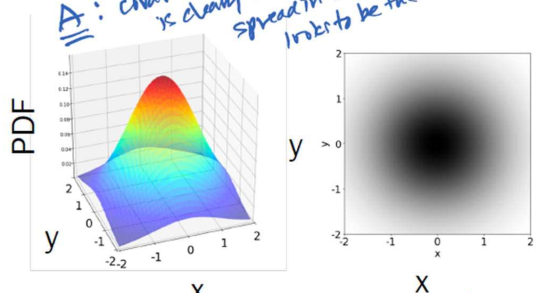
A. $\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$ B. $\begin{bmatrix}1 & 0\\0 & 2\end{bmatrix}$

C. $\begin{bmatrix}1 & 0.5\\0.5 & 1\end{bmatrix}$ D. $\begin{bmatrix}1 & -0.5\\-0.5 & 1\end{bmatrix}$

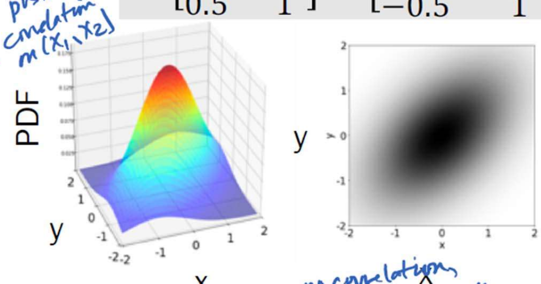
A: covariance is clearly zero, spread in x and y looks to be the same.

C: top-down view suggests positive correlation in (X_1, X_2)

1.



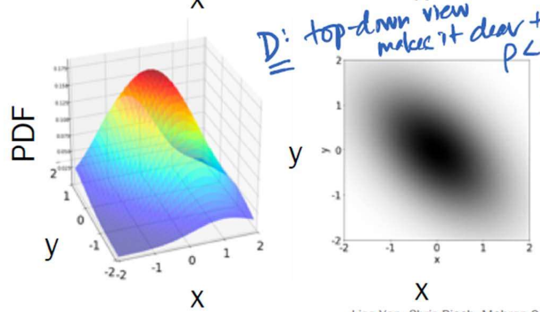
2.



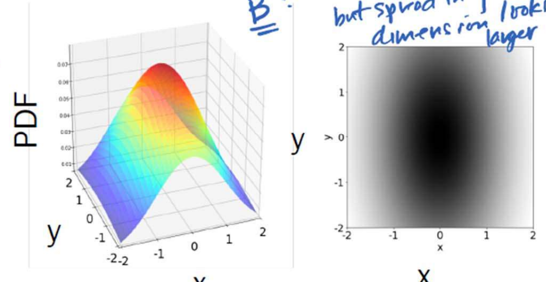
D: top-down view makes it clear that $\rho < 0$, because y decreases as x increases

B: no correlation, but spread in y dimension looks larger

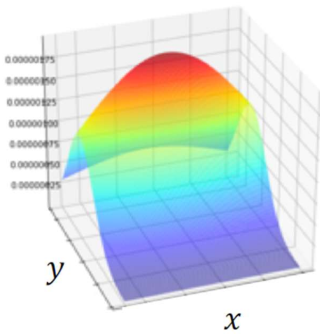
3.



4.



Why are joint PDFs useful?



Independence
2-D support
Joint PDF
Joint CDF
Marginal PDF
(Friday) Conditional PDF

- How 2 continuous RVs vary with each other — $P(X < Y)$, $\mathrm{Cov}(X, Y), \rho(X, Y)$
- How continuous RV is distributed given evidence (more on Friday) — Given $Y = y$, the distribution of X
- How a continuous RV can be decomposed into 2 RVs (or vice versa) — Distribution of $Z = X + Y$ (which is a 1-D RV!)

Sum of Independent Gaussians

Sum of independent Gaussians

$E[X+Y] = E[X]+E[Y] = \mu_1+\mu_2$

$\underbrace{Var(X+Y) = Var(X) + Var(Y)}_{\text{because of independence}} = \sigma_1^2+\sigma_2^2$

$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$

$X \sim \mathcal{N}(\mu_1, \sigma_1^2)$,
$Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$
X, Y independent

$$\Rightarrow \quad X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

(proof left to Wikipedia)

Holds in general case:

$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$
X_i independent for $i = 1, \ldots, n$

$$\Rightarrow \quad \sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Back for another playoffs game

What is the probability that the Warriors win?
How do you model zero-sum games?

$$P(A_W > A_B)$$

This is a probability of an event involving **two** random variables!

We will compute:

$$P(\underbrace{A_W - A_B}_{\text{A sum of Normals!}} > 0)$$

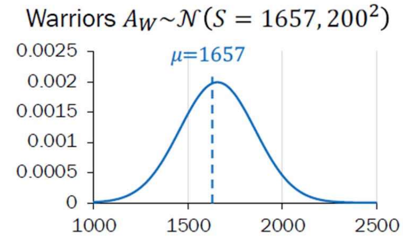
Motivating idea: Zero sum games

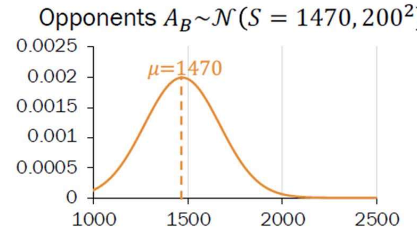
Want: $P(\text{Warriors win}) = P(A_W - A_B > 0)$

Assume A_W, A_B are independent.
Let $D = A_W - A_B$.

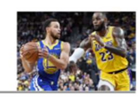
What is the distribution of D?

A. $D \sim \mathcal{N}(1657 - 1470,\ 200^2 - 200^2)$
B. $D \sim \mathcal{N}(1657 - 1470,\ 200^2 + 200^2)$
C. $D \sim \mathcal{N}(1657 + 1470,\ 200^2 + 200^2)$
D. $D \sim \mathcal{N}(1657 + 1470,\ 200^2)$





Motivating idea: Zero sum games



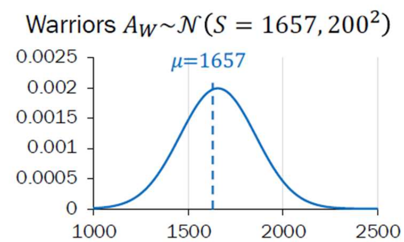
Want: $P(\text{Warriors win}) = P(A_W - A_B > 0)$

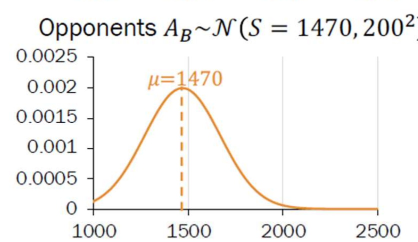
Assume A_W, A_B are independent.
Let $D = A_W - A_B$. } this is really $A_W + (-A_B)$

$-A_B \sim N(-1470, (-1)^2 200^2)$
$\Downarrow$
$N(-1470, 200^2)$

What is the distribution of D?

A. $D \sim \mathcal{N}(1657 - 1470,\ 200^2 - 200^2)$
B. $D \sim \mathcal{N}(1657 - 1470,\ 200^2 + 200^2)$
C. $D \sim \mathcal{N}(1657 + 1470,\ 200^2 + 200^2)$
D. $D \sim \mathcal{N}(1657 + 1470,\ 200^2)$





If $X \sim \mathcal{N}(\mu_1, \sigma^2)$,
then $(-X) \sim \mathcal{N}(-\mu, (-1)^2\sigma^2 = \sigma^2)$.

Motivating idea: Zero sum games

Want: $P(\text{Warriors win}) = P(A_W - A_B > 0)$

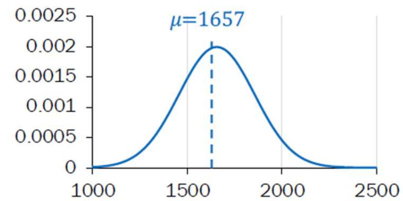
Assume A_W, A_B are independent.
Let $D = A_W - A_B$.

$$D \sim \mathcal{N}(1657 - 1470,\ \ 200^2 + 200^2)$$
$$\sim \mathcal{N}(187, 2 \cdot 200^2) \qquad \sigma \approx 282.842$$

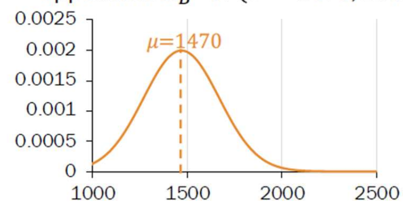
$$P(D > 0) = 1 - F_D(0) = 1 - \Phi\left(\frac{0 - 187}{282.842}\right)$$
$$\approx 0.74574$$

Compare with 0.7488, calculated by sampling!

Warriors $A_W \sim \mathcal{N}(S = 1657, 200^2)$



Opponents $A_B \sim \mathcal{N}(S = 1470, 200^2)$




```
>>> from scipy.stats import norm
>>> 1 - norm(187, 80000 ** 0.5).cdf(0)
0.7457402843526317
>>> 1 - norm(0, 1).cdf(-187 / (80000 ** 0.5))
0.7457402843526317
```


Virus infections

Suppose you are working with the WHO to initiate a response to the onset of a virus. There are two exposed groups:

- G1: 20000 people, each independently infected with $p_1 = 0.1$
- G2: 10000 people, each independently infected with $p_2 = 0.4$

What is $P(\text{people infected} \geq 6100)$? An approximation is okay.

1. Define RVs & state goal

Let A = # infected in G1.
$A \sim \text{Bin}(20000, 0.1)$
B = # infected in G2.
$B \sim \text{Bin}(10000, 0.4)$

Want: $P(A + B \geq 6100)$

Strategy:

A. Sum of independent Binomials — no, $p_1 \neq p_2$:(
B. Sum of independent Poissons — no, because p_1 and p_2 too large
C. Sum of independent Gaussians — yes!!!! :)
D. Sum of independent Exponentials

no, because mean and stdev of B are very different

Virus infections

Suppose you are working with the WHO to initiate a response to the onset of a virus. There are two exposed groups:

- G1: 20000 people, each independently infected with $p_1 = 0.1$
- G2: 10000 people, each independently infected with $p_2 = 0.4$

What is $P(\text{people infected} \geq 6100)$? An approximation is okay.

1. Define RVs & state goal

Let $A = \#$ infected in G1.
$A \sim \text{Bin}(20000, 0.1)$
$B = \#$ infected in G2.
$B \sim \text{Bin}(10000, 0.4)$

Want: $P(A + B \geq 6100)$

2. Approximate as sum of Gaussians

$A \approx X \sim \mathcal{N}(2000, 1800)$ $B \approx Y \sim \mathcal{N}(4000, 2400)$

$P(A + B \geq 6100) \approx P(X + Y \geq 6099.5)$ continuity correction

3. Solve

Virus infections

```
>>> 1 - norm(6000, 4200 ** 0.5).cdf(6099.5)
0.06235282662988528
>>> 1 - norm(0, 1).cdf((6099.5 - 6000)/(4200 ** 0.5))
0.06235282662988528
```

Suppose you are working with the WHO to initiate a response to the onset of a virus. There are two exposed groups:

- G1: 20000 people, each independently infected with $p_1 = 0.1$
- G2: 10000 people, each independently infected with $p_2 = 0.4$

What is $P(\text{people infected} \geq 6100)$? An approximation is okay.

1. Define RVs & state goal

Let $A = \#$ infected in G1.
$A \sim \text{Bin}(20000, 0.1)$
$B = \#$ infected in G2.
$B \sim \text{Bin}(10000, 0.4)$

Want: $P(A + B \geq 6100)$

2. Approximate as sum of Gaussians

$A \approx X \sim \mathcal{N}(2000, 1800)$ $B \approx Y \sim \mathcal{N}(4000, 2400)$

$P(A + B \geq 6100) \approx P(X + Y \geq 6099.5)$ continuity correction

3. Solve

Let $W = X + Y \sim \mathcal{N}(6000, 4200)$

$$P(W \geq 6099.5) = 1 - \Phi\left(\frac{6099.5 - 6000}{\sqrt{4200}}\right)$$

$$\approx 1 - \Phi(1.53531) \approx 0.06235$$

Sum of independent Gaussians

$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$,
$X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$
X_1, X_2 independent

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Is this related to linear transformations of Gaussians?

Recall:

If $Y = aX + b$, then $Y \sim \mathcal{N}(a\mu_X + b, a^2\sigma_X^2)$

Linear transforms vs. independence

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = X + X$. What is the distribution of Y?

- Are both approaches valid?

Independent RVs approach ✖

Let $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$
be independent.
Then $Y = X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$Y = X + X$

$X + X \sim \mathcal{N}(\mu + \mu, \sigma^2 + \sigma^2)$?

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

X is NOT independent of X!

Linear transform approach ✅

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

$Y = 2X$

$Y \sim \mathcal{N}(2\mu, 4\sigma^2)$

For independent $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$,
$aX_1 + bX_2 + c \sim \mathcal{N}(a\mu_1 + b\mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2)$