Conditional Densities - continued

p.91 Bivariate normal

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\underbrace{\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}}_{\text{mean}}, \underbrace{\begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}}_{\text{variance covariance matrix}} \right)$$

$$\sim N_2(\underset{\sim}{\mu}, \Sigma)$$

Joint distribution using mean vector $\underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$,

variance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$ ← covariance

and vector $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$:

Joint

$$f(x_1, x_2) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} (\underset{\sim}{x} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu}) \right\}$$

p is dimension, for 2 variables, p = 2.

Conditional

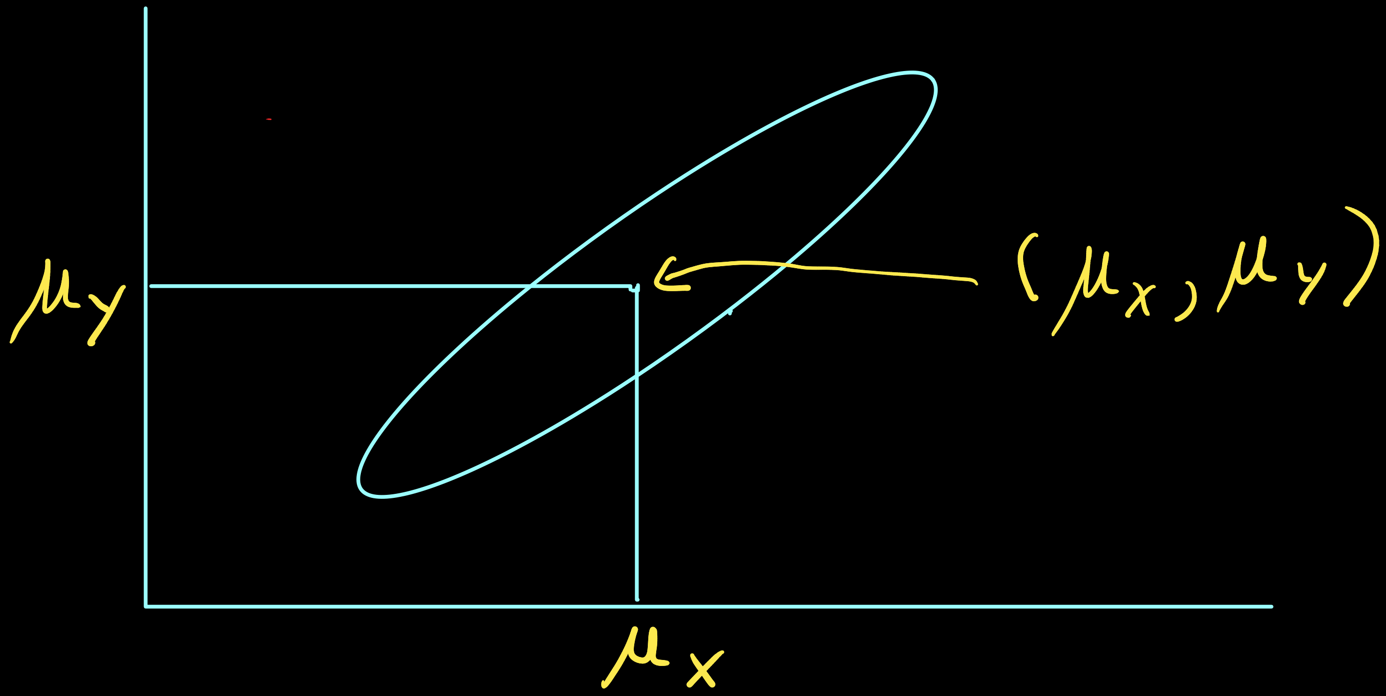
$$f(x_2 | x_1) = \frac{1}{\sqrt{2\pi}\ \sigma_{x_2|x_1}} \exp\left\{ -\frac{1}{2} \left(\frac{x_2 - \left(\mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(x_1 - \mu_1)\right)}{\sigma_{x_2|x_1}} \right)^2 \right\}$$

where $\sigma_{x_2|x_1} = \sqrt{\sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}}$

Distribution of $Y \mid X = x$

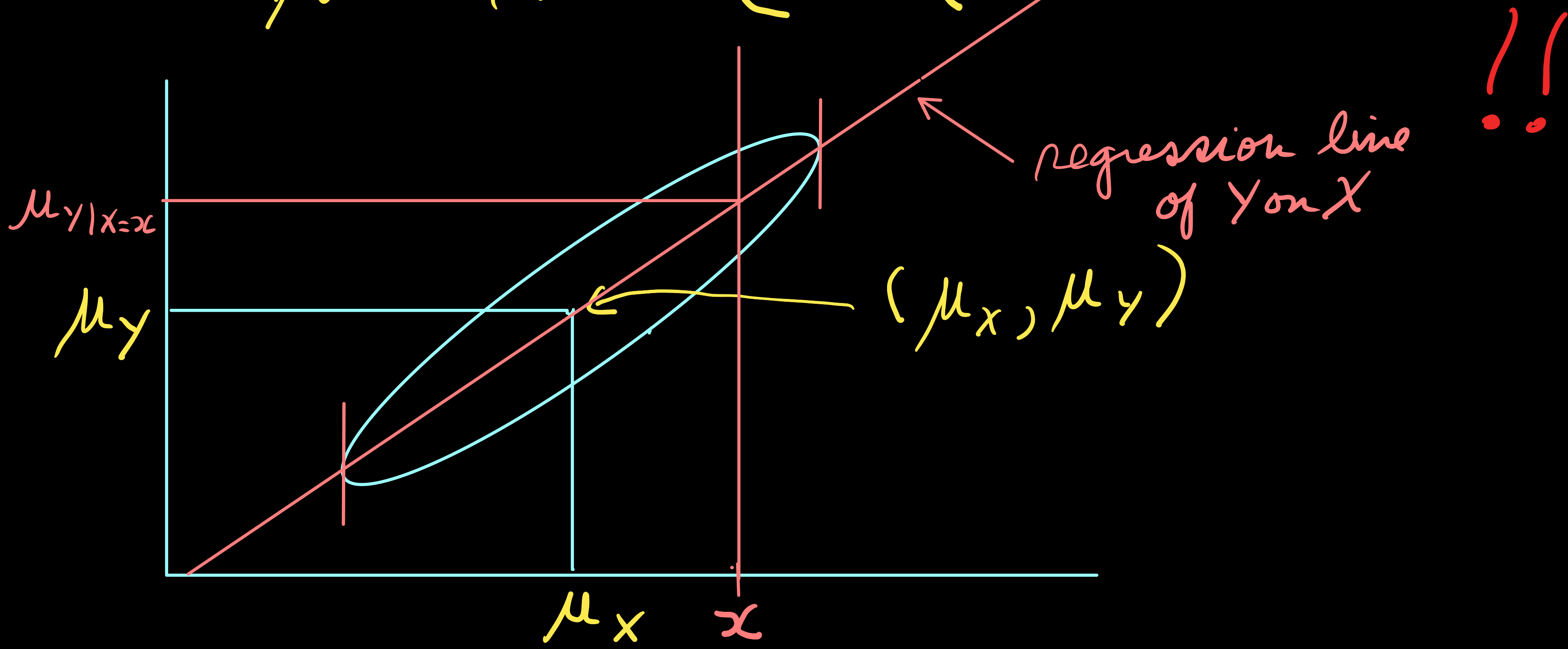
$$f_{Y|X}(y|x) = \frac{1}{\sigma_y\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)}{\sigma_y\sqrt{1-\rho^2}}\right)^2\right\}$$

!!



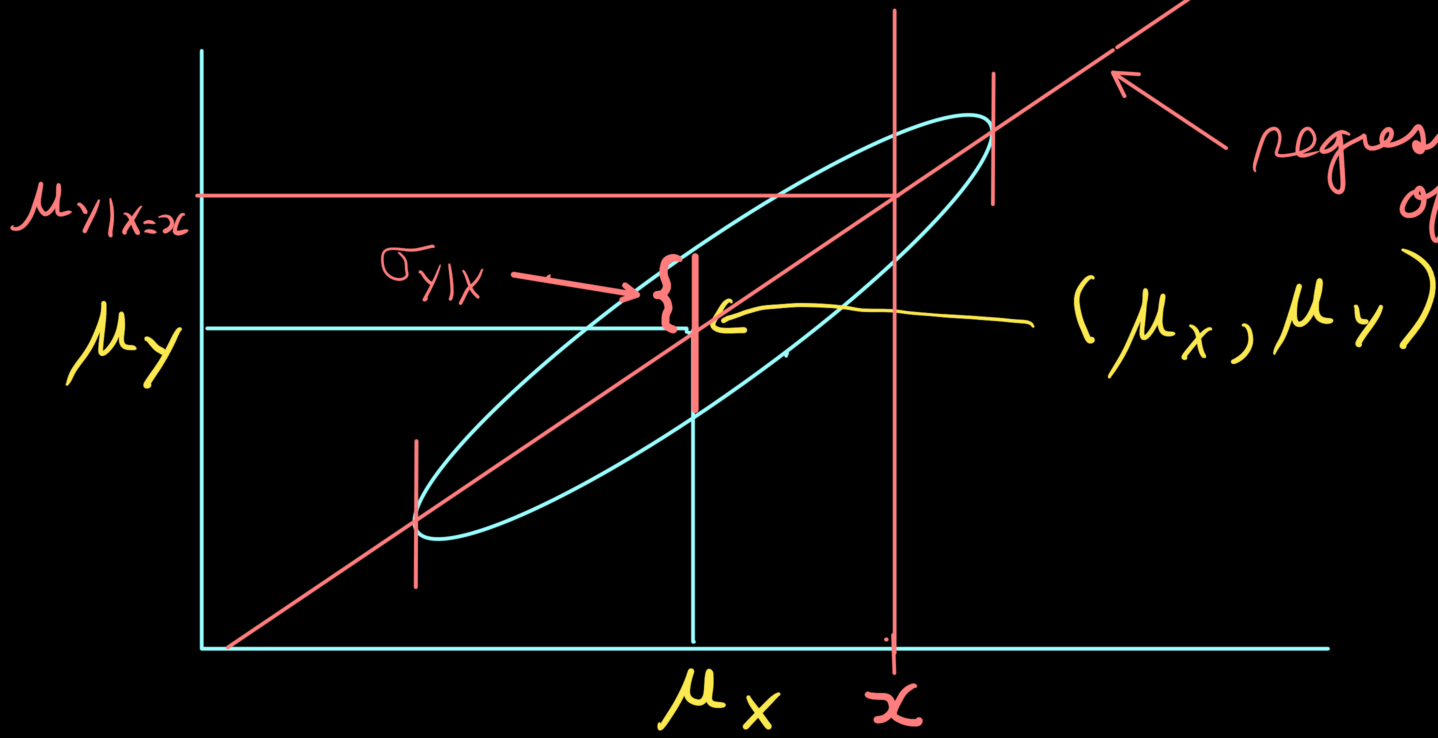
Distribution of $Y \mid X = x$

$$f_{Y|X}(y|x) = \frac{1}{\sigma_Y\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)}{\sigma_Y\sqrt{1-\rho^2}}\right)^2\right\}$$



Distribution of $Y \mid X = x$

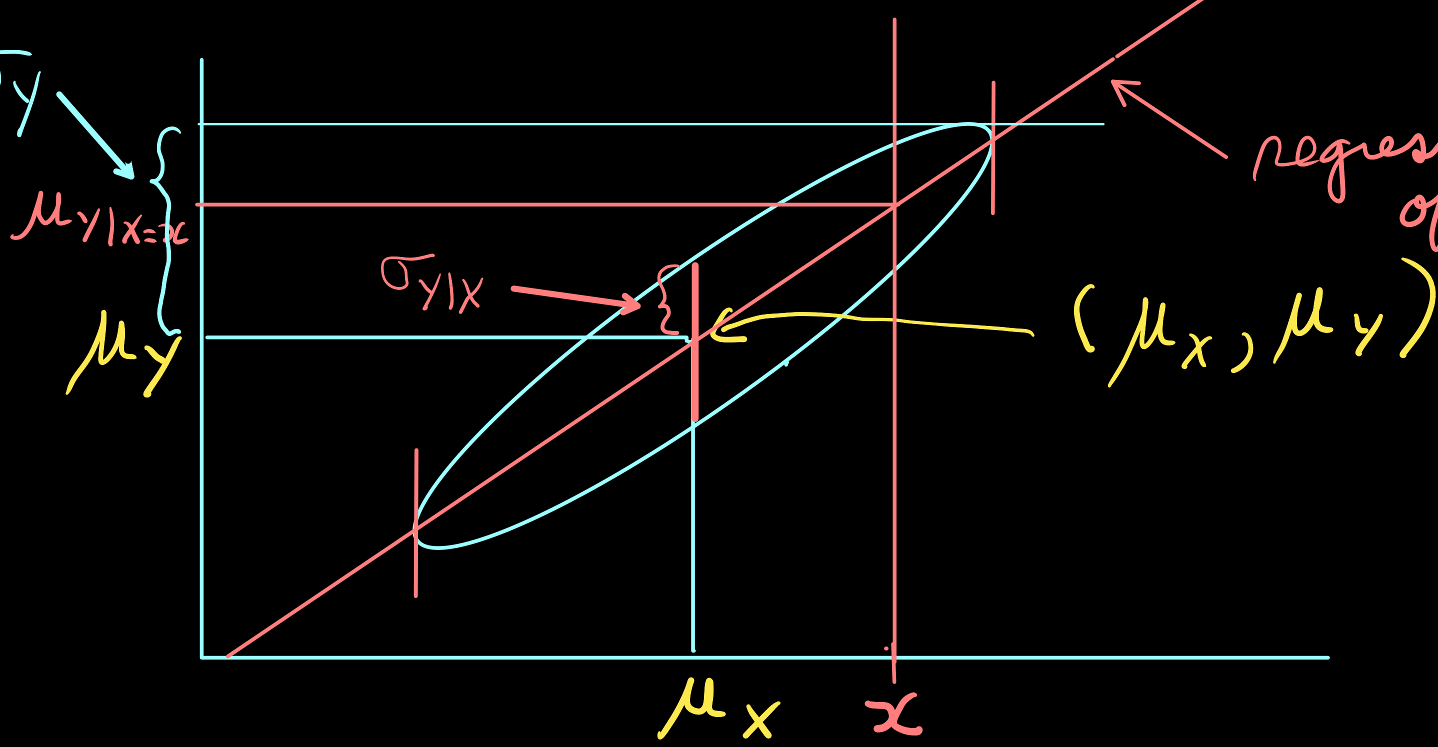
$$f_{Y|X}(y|x) = \frac{1}{\sigma_y\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)}{\sigma_y\sqrt{1-\rho^2}}\right)^2\right\}$$



Distribution of $Y \mid X = x$

$$f_{Y|X}(y|x) = \frac{1}{\sigma_Y\sqrt{2\pi(1-\rho^2)}}\exp\left\{-\frac{1}{2}\left(\frac{y-\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)}{\sigma_Y\sqrt{1-\rho^2}}\right)^2\right\}$$

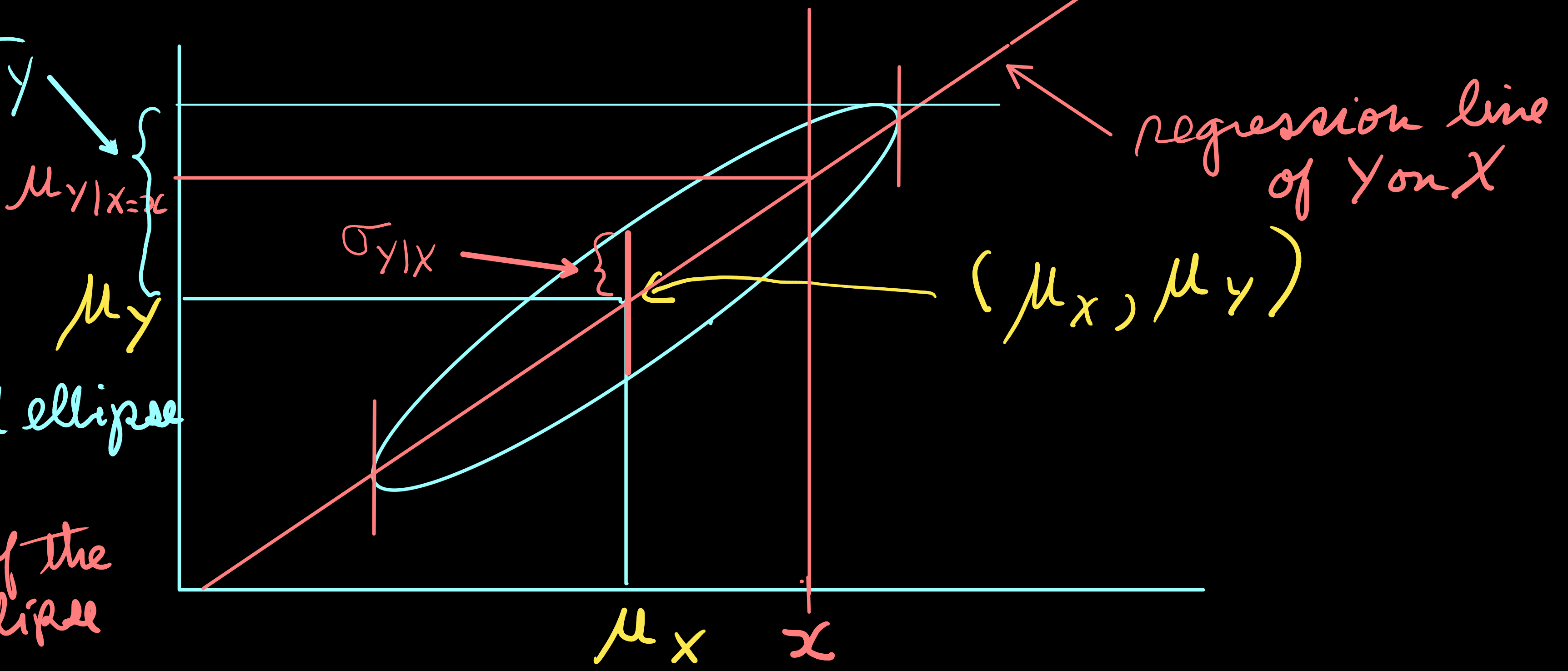
Recall: σ_Y



Distribution of $Y \mid X = x$

$$f_{Y|X}(y|x) = \frac{1}{\sigma_Y\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{1}{2}\left(\frac{y-\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)}{\sigma_Y\sqrt{1-\rho^2}}\right)^2\right\}$$

Recall: σ_Y



σ_Y is the half-shadow of the standard ellipse

$\sigma_{Y|X}$ is the half-slice of the standard ellipse

Generating random variables

Back to p. 63

For continuous CDF's F

Proposition "C"

If X has a continuous strictly monotone CDF, F on an interval, I, then $U = F(X)$ has a $U(0,1)$ distribution

Proof: $Pr(U \leq u) = Pr(F(X) \leq u)$

$$= Pr(X \leq F^{-1}(u))$$

$$= F(F^{-1}(u)) = u$$

This works even if CDF is not monotone BUT NOT if CDF is discrete

Proposition "D"

Let F be continuous, strictly monoton on an interval I

Let $U \sim U(0,1)$. Let $X = F^{-1}(U)$.

Then CDF of X is F.

Proof: $\Pr(X \le x) = \Pr(F^{-1}(U) \le x)$

$= \Pr(U \le F(x)) = F(x)$

Why is this important and useful?

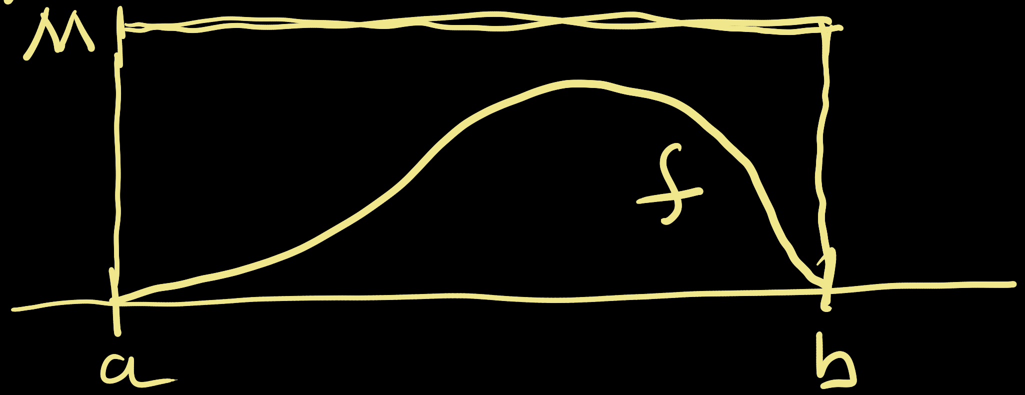
- We know how to generate $U(0,1)$ random variables. So if we can program F, then we can generate random observations of $X = F(U)$

Rejection Method: p. 92

- Suppose X has density f with $f(x) = 0$ outside an interval $[a, b]$
- Could use $X = F^{-1}(U)$ if F^{-1} is easy to program.
- But sometimes, we can program f, but F or F^{-1} very hard.
- If f has a maximum ~~on~~ on $[a, b]$ we can use the rejection method.

Simple algorithm

0) Choose $M \geq \max_x f(x)$



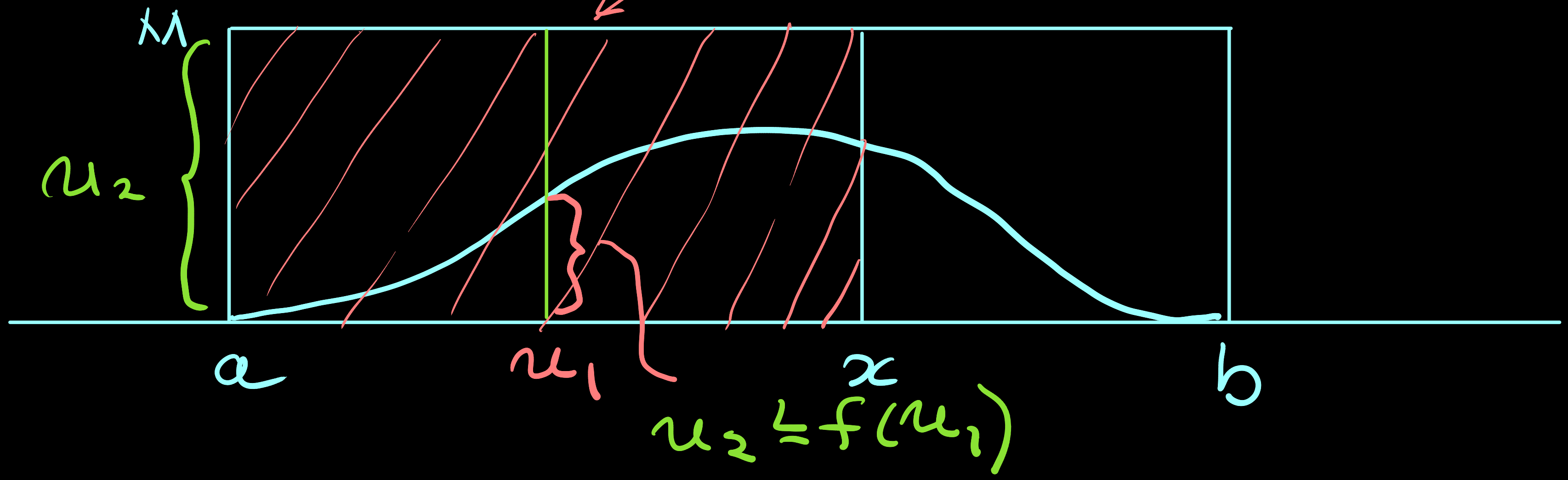
1) Let $U_1 \sim \text{Uniform}(a, b)$
2) Let $U_2 \sim \text{Uniform}(0, M)$
3) Let $X = \begin{cases} U_1 & \text{if } U_2 \leq f(U_1) \\ \text{go back to (1) otherwise} \end{cases}$

repeat until you have enough X's

If $U \sim U(0,1)$
then $Z = (b-a)U + a$
$\sim \text{Uniform}(a, b)$

Let's check whether this works:

$$\Pr(X \leq x) = \Pr(\underbrace{U_1 \leq x} \mid U_2 \leq f(U_1))$$



$$\Pr(X \in x \pm dx) = \frac{\Pr(U_1 \in x \pm dx \cap U_2 \leq f(U_1))}{\Pr(U_2 \leq f(U_1))}$$

$$= \frac{\Pr(U_2 \leq f(U_1) \mid U_1 \in x \pm dx)\Pr(U_1 \in x \pm dx)}{\Pr(U_2 \leq f(U_1))}$$

$$= \frac{f(x)}{M} \times \frac{dx}{b-a} \Big/ \left(\int_a^b f(x)\,dx \Big/ M(b-a) \right)$$

$$= f(x) / 1 = f(x)$$

!!!

Example E Bayesian Inference

20 sided die Y = # on face

If fair $Pr(Y \le 10) = 1/2$

But maybe not.

Let $Pr(Y \le 10) = \theta$, $0 \le \theta \le 1$

We don't know θ!

Baysian approach:

- Use a probability distribution to represent your uncertainty about θ.
- Tough act. But let's pretend

$$\theta \sim U(0,1) \quad \text{is reasonable.}$$

"prior distribution"

- Get data: Toss die n times
- Let x be # of times $Y \le 10$

What do we know about θ now?

Distribution of θ given $X = x$

Posterior distribution

$$p(x\mid\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

$$p(x,\theta) = p(x\mid\theta)\,p(\theta)$$

$$= \binom{n}{x}\theta^x(1-\theta)^{n-x} \times 1, \quad 0<\theta<1$$

We want $p(\theta\mid x)$

so need $p(x) = \int_0^1 \binom{n}{x}\theta^x(1-\theta)^{n-x}\,d\theta$

$$= \ldots = \frac{1}{n+1}, \quad x=0,\ldots,n$$

So $p(\theta|x) = (n+1)\binom{n}{x}\theta^{x}(1-\theta)^{n-x} \quad 0 < \theta < 1$

This is a Beta$(x+1, n-x+1)$

and has mean $\dfrac{x+1}{n+2}$

The "frequentist" estimate of θ is $\dfrac{x}{n}$

Very close to $\dfrac{x+1}{n+2}$ if n large

and x not close to 0 or n.