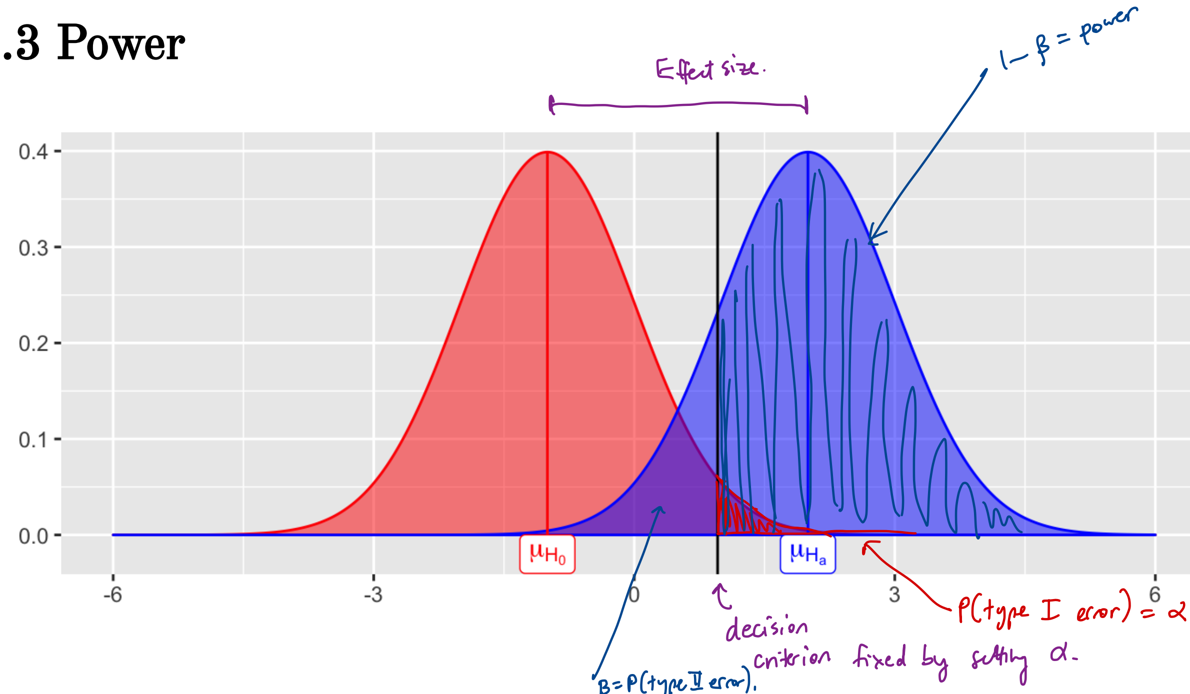
2.3 Power



Consider a hypothesis test about the parameter θ:

$$H_0 : \theta = \theta_0$$
$$H_a : \theta > \theta_0$$

We let $\beta = P(\text{fail to reject } H_0 | H_0 \text{ is false}) = P(\text{Type II error})$, then Power $=$ $P(\text{reject } H_0 | H_0 \text{ is false}) = 1 - \beta$.

Power depends on the distance between the hypothesized value of the parameter θ_0 and the actual value θ_1, so we can write $1 - \beta(\theta_1)$.

the effect size.

Why is power important?

1. If we have multiple statistical testing methods for the same hypothesis (similar size), choose test that is the most powerful.

2. If we are going to spend time/money to run an experiment, need to check beforehand that the study is powerful enough to detect an effect.

For a few simple cases, you can derive a closed form expression of power.

All others: use Monte Carlo methods to estimate power.

known.

Example 2.4 Consider a one-sample z-test. Sample $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

unknown

To test $H_0: \mu = \mu_0$, $H_a: \mu > \mu_0$ using statistic $z^* = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$,

we reject H_0 if $z^* > z_{1-\alpha}$ critical value $1-\alpha$ quantile of $N(0,1)$.

If $\mu_0 = 5$ (hypothesized value) but true mean $\mu_1 = 6$

What is the probability of correctly rejecting $H_0: \mu = 5$? This is Power!

Effect size: $\mu_1 - \mu_0 = 6 - 5 = 1$. If effect size is 10, our test will have more power (easier to detect the truth).

For the z-test, we can derive power (Chihara & Hesterberg p. 229-230).

$$1 - \beta = P(\text{reject } H_0 \mid H_0 \text{ is false})$$
$$= P\Big(z^* > \underbrace{z_{1-\alpha} - \frac{(\boxed{\mu_1 - \mu_0})}{\sigma/\sqrt{n}}}_{\text{smallest } z \text{ where you will reject } H_0}\Big)$$

So power is a function of

1. Significance level: as $\alpha \uparrow$, power $\uparrow$ (trade-off between type I and type II error).

2. Effect size: $\mu_1 - \mu_0$ as effect size $\uparrow$, power $\uparrow$.

3. Sample size: as $n\uparrow$, power $\uparrow$

4. Variance: as variance $\downarrow$, power $\uparrow$ (no control over this in practice).

Notes: ① as power $= 1-\beta \uparrow$, $P(\text{type I error}) = \alpha \uparrow$. For fixed n, σ, & $\mu_1 - \mu_0$, the only way to increase power is to $\uparrow \alpha$

② Only way to simultaneously $\uparrow$ power & $\downarrow \alpha$, must $\uparrow n$. (fixed effect size).

2.4 MC Estimator of $1 - \beta$

Assume $X_1, \dots, X_n \sim F(\theta_0)$ (i.e., assume H_0 is true).

Then, we have the following hypothesis test –

$$H_0 : \theta = \theta_0$$
$$H_a : \theta > \theta_0$$

and the statistics T^*, which is a test statistic computed from data. Then we **reject** H_0 if $T^* >$ the critical value from the distribution of the test statistic.

This leads to the following algorithm to estimate the power of the test $(1 - \beta)$

1. Set up hypothesis test (pick T^*, determine dsn of T^*).

new * 2. Select value of alternative $\theta_1 > \theta_0$

3. set n, other parameter values, and α

4. For each $j = 1, \dots, m$:

   * a) Sample $X_1^{(j)}, \dots, X_n^{(j)}$ from model under the alternative hypothesis $\theta = \theta_1$.

   b) Compute T_j based on data from a)

   c) $y_j = \mathbb{I}(\underline{\text{reject } H_0 \text{ based on } T_j})$
   $= \mathbb{I}(T_j > \text{crit value})$

5. Compute $1 - \hat{\beta} = \frac{1}{m} \sum_{j=1}^{m} y_j$ (i.e. $\frac{\text{count of correct answers}}{\text{\# times we tried}}$)

Your Turn

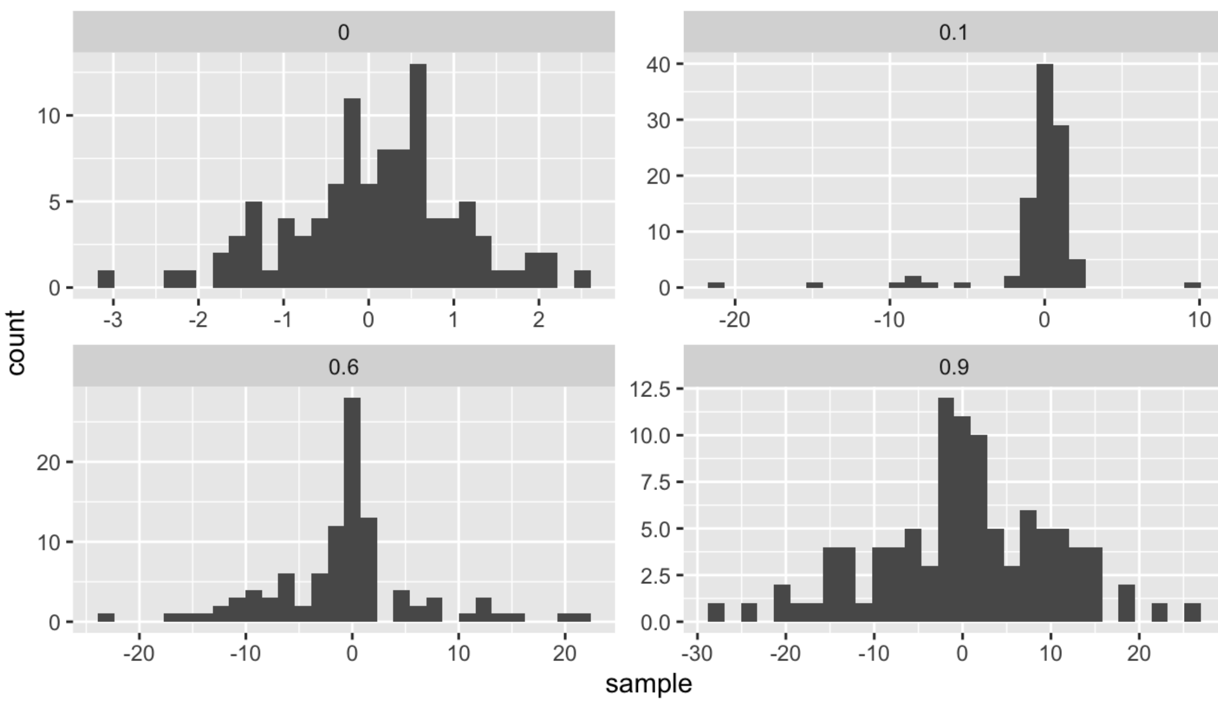
Consider data generated from the following mixture distribution:

$$f(x) = (1-\epsilon)f_1(x) + \epsilon f_2(x), \quad x \in \mathbb{R}$$

where f_1 is the pdf of a $N(0,1)$ distribution, f_2 is the pdf of a $N(0,100)$ distribution, and $\epsilon \in [0,1]$.

```
r_noisy_normal <- function(n, epsilon) {
  z <- rbinom(n, 1, 1 - epsilon)
  z*rnorm(n, 0, 1) + (1 - z)*rnorm(n, 0, 10)
}

n <- 100
data.frame(e = 0, sample = r_noisy_normal(n, 0)) %>%
  rbind(data.frame(e = 0.1, sample = r_noisy_normal(n, 0.1))) %>%
  rbind(data.frame(e = 0.6, sample = r_noisy_normal(n, 0.6))) %>%
  rbind(data.frame(e = 0.9, sample = r_noisy_normal(n, 0.9))) %>%
  ggplot() +
  geom_histogram(aes(sample)) +
  facet_wrap(.~e, scales = "free")
```



We will compare the power of various tests of normality. Let F_X be the distribution of a random variable X. We will consider the following hypothesis test,

i.e. H_0 says X is Normally distributed and H_a says it isn't.

$$H_0 : F_x \in N \qquad \text{vs.} \qquad H_a : F_x \notin N,$$

where N denotes the family of univariate Normal distributions.

Recall Pearson's moment coefficient of skewness (See Example 2.2). and corresponding skewness test.

We will compare Monte Carlo estimates of power for different levels of contamination ($0 \leq \epsilon \leq 1$). We will use $\alpha = 0.1$, $n = 100$, and $m = 100$.

$H_0 : \sqrt{\beta_1} = 0$

$H_a : \sqrt{\beta_1} \neq 0.$

```r
# skewness statistic function
skew <- function(x) {
  xbar <- mean(x)
  num <- mean((x - xbar)^3)
  denom <- mean((x - xbar)^2)
  num/denom^1.5
}

# setup for MC
alpha <- .1
n <- 100
m <- 100
epsilon <- seq(0, 1, length.out = 200)
var_sqrt_b1 <- 6*(n - 2)/((n + 1)*(n + 3)) # adjusted variance for
  skewness test
crit_val <- qnorm(1 - alpha/2, 0, sqrt(var_sqrt_b1)) #crit value for
  the test
empirical_pwr <- rep(NA, length(epsilon)) #storage

# estimate power for each value of epsilon
for(j in 1:length(epsilon)) {
  # perform MC to estimate empirical power
  ## Your turn

}

## store empirical se
empirical_se <- "Your Turn: fill this in"

## plot results --
## x axis = epsilon values
## y axis = empirical power
## use lines + add band of estimate +/- se
```

$\sqrt{\hat{p}(1-\hat{p})/m}$

Compare the power with $n = 100$ to the power with $n = 10$. Make a plot to compare the two for many values of ϵ.

this is similar to effect size.

- for what levels of contamination (ε) will we have power ≥ 0.8

Recall power depends on

1. level of test
2. sample size
3. effect size
4. variance.