

Week 01 Practice Problems (Optional)

2026-08-26

```
# this is setup for Julia code in the solutions
using CairoMakie
using Distributions
using Printf
using Random
```

You don't need to finish these by Wednesday of week 1, but take a look at these problems to get a sense of what you will find on exams.

Repeated exceedances in a fixed window

A flood level is exceeded on average once every ten years. What is the probability it is exceeded at least twice in the next twenty years? State the assumption your calculation makes about different years.

i Answer

Let Y be the number of exceedances in twenty years, so $Y \sim \text{Binomial}(20, p)$ with $p = 1/10$. Then

$$\begin{aligned} P(Y \geq 2) &= 1 - P(Y = 0) - P(Y = 1) \\ &= 1 - 0.9^{20} - 20(0.1)(0.9)^{19} \end{aligned}$$

which we can compute in Julia as

```
dist = Binomial(20, 1/10)
@printf("P(Y >= 2) = %.3g", 1 - pdf(dist, 0) - pdf(dist, 1))
```

```
P(Y >= 2) = 0.608
```

ccdf does the same in one call.

```
@printf("P(Y >= 2) = %.3g", ccdf(dist, 1))
```

```
P(Y >= 2) = 0.608
```

The binomial step assumes years are independent and identically distributed.

Design life and return period

A temporary works must be protected so that the chance of being overtopped at least once during its four-year life is no more than 10 percent.

- What return period must the design event have?
- Suppose someone instead designs to the twenty-year event. What is the actual chance of overtopping over the four-year life, and does it meet the specification?

i Answer

Part a. Set the four-year risk to 0.10 and solve for the annual exceedance probability p , then take the return period as $T = 1/p$.

$$1 - (1 - p)^4 = 0.10$$
$$p = 1 - 0.90^{1/4}$$
$$T = 1/p$$

```
p = 1 - 0.90^(1 / 4)
T = 1 / p
@printf("p = %.3g, T = %.3g years", p, T)
```

```
p = 0.026, T = 38.5 years
```

About a 39-year event. The specification caps the risk, so Naghettini (2017) rounds the design target up to the 50-year flood. Rounding to a shorter return period would raise the four-year risk above the 10 percent allowed.

Part b. The four-year risk from the twenty-year event, $p = 0.05$:

```
@printf("4-year risk = %.3g", 1 - 0.95^4)
```

```
4-year risk = 0.185
```

That is nearly twice the risk the specification allows, so the twenty-year event does not meet it. A return period is an average waiting time, and a design life of L years accumulates risk as $1 - (1 - p)^L$.

Nonstationary hazard behind a fixed label

The Federal Flood Risk Management Standard state of the science report gives a 43.17 percent chance of at least one flood of a stated height at Grand Isle, Louisiana over the next 30 years (Honeycutt et al., 2023, p. 81).

- What constant annual exceedance probability would produce that same 30-year chance?

- b. Under stationary conditions, what is the chance of at least one 1-percent-annual-chance flood in 30 years?
- c. Compare the two. The stated flood is still described as the 1-percent flood. What is that description now doing?

i Answer

Part a. Invert the same relation over 30 years:

$$p = 1 - (1 - 0.4317)^{1/30}$$

```
@printf("p = %.3g", 1 - (1 - 0.4317)^(1 / 30))
```

```
p = 0.0187
```

Part b. Hold $p = 0.01$ fixed:

```
@printf("30-year risk = %.3g", 1 - 0.99^30)
```

```
30-year risk = 0.26
```

Part c. The annual probability has nearly doubled.

Part a assumes one constant p across all 30 years, so what it returns is the constant annual probability that matches a hazard which is actually rising. The label “1-percent flood” names a water level. It no longer states the annual chance of reaching that level.

Plotting positions and the reach of a record

```
Random.seed!(543)
record = rand(GeneralizedExtremeValue(4.0, 0.8, 0.15), 40)
sorted = sort(record; rev=true)
m = length(sorted)
T_hat = (m + 1) ./ (1:m)

fig = Figure(; size=(700, 400))
ax = Axis(
    fig[1, 1];
    xscale=log10,
    xlabel="return period T (years)",
    ylabel="annual maximum water level (ft)",
    xticks=([1, 2, 5, 10, 20, 40], string.([1, 2, 5, 10, 20, 40])),
)
scatter!(ax, T_hat, sorted; color=:steelblue)
fig
```

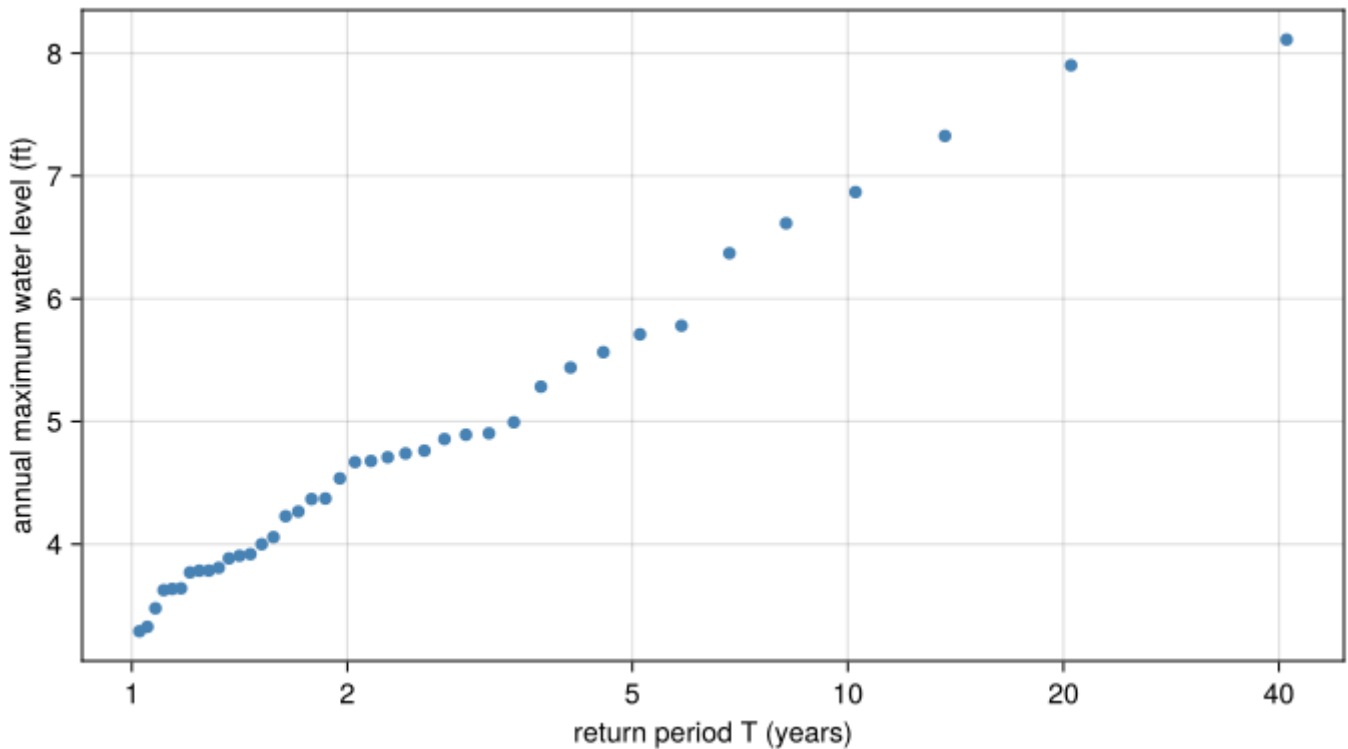


Figure 1: Forty simulated annual maxima, ranked largest to smallest and placed at Weibull plotting positions

Figure 1 is one simulated record of forty annual maxima, ranked from largest to smallest and placed at the Weibull plotting position $\hat{p}_i = i/(m + 1)$, with $\hat{T}_i = 1/\hat{p}_i$.

- What return period does the largest value in the record receive, and what sets that number?
- Read the 10-year level off the plot. Say which coordinate of that point was observed and which was computed.
- A colleague asks you for the 200-year level. What can this plot give them?

i Answer

Part a. Rank $i = 1$ gives

$$\hat{p}_1 = \frac{1}{m+1}$$
$$\hat{T}_1 = m+1$$

```
@printf("m = %d years,  $\hat{T}$  at rank 1 = %.3g years", m, T_hat[1])
```

```
m = 40 years,  $\hat{T}$  at rank 1 = 41 years
```

The length of the record sets it. Nothing about the largest value itself enters, so a different record of the same length puts a different level at the same $\hat{T}$.

Part b.

```
j = argmin(abs.(T_hat .- 10))  
@printf("nearest point:  $\hat{T}$  = %.3g years, level = %.3g ft", T_hat[j], sorted[j])
```

```
nearest point:  $\hat{T}$  = 10.2 years, level = 6.87 ft
```

The level is the observation. The return period is computed from that value's rank, so a different plotting-position formula moves the point sideways without touching the data.

Part c. Nothing. The largest $\hat{T}$ on the plot is $m+1$, so the method cannot report a level for any return period past the length of the record. That ceiling comes from the Weibull formula $i/(m+1)$; other plotting positions exist and have been hotly debated. Reaching 200 years means fitting a distribution and extrapolating, which is module 2.

All three parts assume the forty years are independent draws from one distribution.

Sample size for a target relative error

You want to estimate the probability of the 50-year event by simulating years and counting exceedances, and you want the answer good to about 10 percent of its own size. Roughly how many years must you simulate? Use Wednesday's result, $n = (CV/\varepsilon)^2$, where ε is the standard error as a fraction of the answer. Then check it against the cruder rule that the relative error is about $1/\sqrt{np}$, where p is the probability you are estimating and n is the number of simulated years.

i Answer

The quantity being averaged is the indicator of exceedance, which is Bernoulli with $p = 1/50$, so its coefficient of variation is fixed by p alone.

$$\begin{aligned} CV &= \frac{\sqrt{p(1-p)}}{p} \\ &= \sqrt{\frac{1-p}{p}} \\ &= \sqrt{T-1} \\ n &= \left(\frac{CV}{\varepsilon}\right)^2 = \frac{T-1}{\varepsilon^2} \end{aligned}$$

```
T = 50
ε = 0.10
n = (T - 1) / ε^2
@printf("n = %.0f years, %.0f exceedances", n, n / T)
```

```
n = 4900 years, 98 exceedances
```

The cruder rule drops the $(1 - p)$ factor and gives $n = T/\varepsilon^2$:

```
@printf("n = %.0f years", T / ε^2)
```

```
n = 5000 years
```

Both steps assume each simulated year is an independent draw. Accuracy is set by the expected number of exceedances rather than by n . The two answers differ by 2 percent, which is below the precision this calculation is used at.

Mode, mean, and spread of the waiting time

Let the wait for the next 100-year flood be the number of years until the first exceedance, counting the year it happens.

- What is the most likely wait?
- What is the mean wait?
- What is the standard deviation of the wait?

i Answer

The wait τ is geometric with $p = 1/100$, so $P(\tau = k) = (1 - p)^{k-1} p$.

$$P(\tau = k + 1)/P(\tau = k) = 1 - p$$

$$E[\tau] = 1/p$$

$$\text{Var}[\tau] = (1 - p)/p^2$$

$$\text{SD}[\tau] = \sqrt{1 - p} / p$$

Part a. The ratio is below one at every k , so the pmf is decreasing and the mode is one year.

Parts b and c.

```
p = 1 / 100
@printf("mean = %.3g years, sd = %.3g years", 1 / p, sqrt(1 - p) / p)
```

```
mean = 100 years, sd = 99.5 years
```

The geometric form assumes years are independent with the same p . The most likely wait is one year and the mean wait is 100 years, and the standard deviation is nearly as large as the mean.

Bibliography

- Honeycutt, M., Marcy, D., Kinsman, N., Clark, E., Mason, R., Hecht, J., et al. (2023). *Federal Flood Risk Management Standard Climate-Informed Science Approach (CISA) State of the Science Report*. Retrieved from <https://www.whitehouse.gov/wp-content/uploads/2023/03/Federal-Flood-Risk-Management-Standard-Climate-Informed-Science-Approach-CISA-State-of-the-Science-Report.pdf>
- Naghattini, M. (Ed.). (2017). *Fundamentals of Statistical Hydrology*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-43561-9>