8. Exercises on Statistical models

A Parametric Model for Rock Samples

2/2 points (graded)
You are testing out a new scale that measures weights. To do so, you collect a particular rock and take n measurements, using the same rock each time. Let X_i denote the ith measurement a particular rock.

Based on prior knowledge, you expect your data (*i.e.* statistical experiment) $X_1, \ldots, X_n$ to consist of **i.i.d.** (independent and identically distributed) samples from a Gaussian with unknown mean $\mu > 0$. The scale that you are using to weigh the sample comes with a guarantee from the manufacturer that the variance of your data set will be 0.23. Given this information, your goal is to write down a statistical model $(E, \{P_\theta\}_{\theta\in\Theta})$ for this statistical experiment.

Which of the following is (are)

(1) valid statistical model(s)?

(Choose all that apply.)

- [x] $\left((-\infty,\infty), \{N(\mu, 0.23)\}_{\mu>0}\right)$ ✔
- [x] $\left((-\infty,\infty), \{N(\mu, 0.23)\}_{\mu\in\mathbb{R}}\right)$ ✔
- [x] $\left((-\infty,\infty), \{N(\mu, \sigma^2)\}_{\mu,\sigma^2>0}\right)$ ✔
- [x] $\left((-\infty,\infty), \{N(\mu, \sigma^2)\}_{\mu\in\mathbb{R},\sigma^2>0}\right)$ ✔

✔

(2) the statistical model that **best** incorporates all known information?

- [x] $\left((-\infty,\infty), \{N(\mu, 0.23)\}_{\mu>0}\right)$ ✔
- [ ] $\left((-\infty,\infty), \{N(\mu, 0.23)\}_{\mu\in\mathbb{R}}\right)$
- [ ] $\left((-\infty,\infty), \{N(\mu, \sigma^2)\}_{\mu,\sigma^2>0}\right)$
- [ ] $\left((-\infty,\infty), \{N(\mu, \sigma^2)\}_{\mu\in\mathbb{R},\sigma^2>0}\right)$

✔

Solution:

1. All of the above choices are valid statistical models. The sample space of a Gaussian is $(-\infty,\infty)$. In the first and second choice, $\left((-\infty,\infty), \{N(\mu, 0.23)\}_{\mu>0}\right)$ and $\left((-\infty,\infty), \{N(\mu, 0.23)\}_{\mu\in\mathbb{R}}\right)$ use the mean μ to parametrize the Gaussian. In the third and fourth choice, $\left((-\infty,\infty), \{N(\mu, \sigma^2)\}_{\mu,\sigma^2>0}\right)$ and $\left((-\infty,\infty), \{N(\mu, \sigma^2)\}_{\mu\in\mathbb{R},\sigma^2>0}\right)$, parametrize the Gaussian distribution by the mean and variance. Since these choices restrict $\sigma^2 > 0$, both are valid statistical models.

2.

For the purposes of modeling, in general it is best to choose the statistical model that incorporates all known information about the sample. Usually this reduces the amount of unknowns in the model or the size of the parameter space. Since we are given that the data is Gaussian, the variance is 0.23, and the mean μ, is positive, it makes sense to incorporate this information into the model. Note that choice 1 $\left((-\infty,\infty), \{N(\mu, 0.23)\}_{\mu>0}\right)$ uses everything that we are given in the problem set-up, so it is likely the best choice of statistical model in this scenario.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Parametric vs. Nonparametric Models

1/1 point (graded)

A statistical model $(E, \{P_\theta\}_{\theta\in\Theta})$ is **parametric** if all parameters $\theta \in \Theta$ can be specified by a **finite** number of unknowns. Equivalently, this means that Θ is a subset of $\mathbb{R}^m$. In particular, if $\Theta \subset \mathbb{R}^m$, then P_θ is uniquely specified by the m entries of the vector θ.

Which of the following statistical models are parametric?
(Choose all that apply.)

- [ ] $E = \{x \in \mathbb{Z} : x \geq 0\}$;
  $\{P_\theta\}_{\theta\in\Theta}$ is the set of all probability distributions with the sample space $\{x \in \mathbb{Z} : x \geq 0\}$.

- [x] $E = \{0, 1\}$;
  $\{P_\theta\}_{\theta\in[0,1]} = \{\mathrm{Ber}(\theta)\}_{\theta\in[0,1]}$. ✔

- [x] $E = (-\infty, \infty)$;
$\{P_{\sigma^2}\}_{\sigma^2 \in (0,\infty)}$ is the set of all centered (mean 0) Gaussian distributions $\mathcal{N}(0, \sigma^2)$ where $\sigma^2 > 0$. ✔

- [x] $E = \{1, 2, 3, 4\}$;
$\{P_{(p_1,p_2,p_3,p_4)}\}_{(p_1,p_2,p_3,p_4)\in S}$ is defined in terms of
  - S: the set of all $(p_1, p_2, p_3, p_4) \in \mathbb{R}^4$ such that $0 \le p_i \le 1$ for all $i = 1, \ldots, 4$ and $\sum_{i=1}^{4} p_i = 1$;
  - $P_{(p_1,p_2,p_3,p_4)}$: the distribution defined by setting the probability of outcome i to be p_i.

  ✔

- [x] $E = (-\infty, \infty)$;
$\{P_{(\mu,\sigma^2)}\}_{(\mu,\sigma^2)\in\mathbb{R}\times(0,\infty)}$ is the set of all Gaussian distributions $\mathcal{N}(\mu, \sigma^2)$ where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. ✔

- [x] $E = (0, \infty)$;
$\{P_\theta\}_{\theta\in[0,\infty)} = \{\mathcal{U}([0, \theta])\}_{\theta\in(0,\infty)}$ is the set of all uniform distributions on the interval $[0, \theta]$ with $\theta > 0$. ✔

- [ ] $E = [0, 1]$;
$\{P_\theta\}_{\theta\in\Theta}$ is the set of all probability distributions given by a probability density function $f : \mathbb{R} \to \mathbb{R}$ with f continuous and $\int_0^1 f(x)\, dx = 1$.

✔

Solution:

- "$E = \{x \in \mathbb{Z} : x \geq 0\}$ and $\{P_\theta\}_{\theta \in \Theta}$ is the set of all probability distributions with sample space $\{x \in \mathbb{Z} : x \geq 0\}$.", specifying the distribution requires us to know the probability of the outcome i for all $i \in \mathbb{Z}$ such that $i \geq 0$. An infinite amount of information (or unknowns) is required, so this statistical model is non-parametric.

- $E = \{0, 1\}$ and $\{P_\theta\}_{\theta \in [0,1]} = \{\mathrm{Ber}\,(\theta)\}_{\theta \in [0,1]}$.",
"$E = (-\infty, \infty)$ and $\{P_{\sigma^2}\}_{\sigma^2 \in (0,\infty)}$ is the set of all centered (mean 0) Gaussian distributions...", and
"$E = (-\infty, \infty)$ and $\{P_\theta\}_{\theta \in [0,\infty)} = \{\mathcal{U}\,([0, \theta])\}_{\theta \in [0,\infty)}$...," respectively, all require just a single unknown to specify the distribution. These models are parametric.

- The choice "$E = \{1, 2, 3, 4\}$ and $\{P_{(p_1,p_2,p_3,p_4)}\}_{(p_1,p_2,p_3,p_4) \in S}$..." requires three unknowns to specify the distribution (once p_1, p_2, and p_3 are specified, p_4 is uniquely determined). This model is parametric. It would remain parametric even if one said, “there are four unknowns, p_1, p_2, p_3, p_4".

- The choice "$E = (-\infty, \infty)$ and $\{P_{(\mu,\sigma^2)}\}_{(\mu,\sigma^2) \in \mathbb{R} \times (0,\infty)}$ is the set of all Gaussian distributions $N\,(\mu, \sigma^2)$..." requires only the specification of the mean and variance, so it is also parametric.

- Similarly, in the last choice "$E = [0, 1]$ and $\{P_\theta\}_{\theta \in \Theta}$ is the set of all probability distributions given by a probability density function $f : \mathbb{R} \to \mathbb{R}$...", the space of continuous density functions cannot be specified by a finite amount of information; you would need to know the values of the function on an infinite subset of $[0, 1]$ to be able to uniquely determine it. Hence, this statistical model is also non-parametric.

Submit

You have used 1 of 7 attempts

Answers are displayed within the problem

Statistical Model for a Censored Exponential

1/1 point (graded)

Let X denote an exponential random variable with unknown parameter $\lambda > 0$. Let $Y = \mathcal{I}(X > 5)$, the indicator that X is larger than 5.

Recall the definition of the indicator function here is

$$\mathcal{I}(X > 5) = \begin{cases} 1 & \text{if } X > 5 \\ 0 & \text{if } X \leq 5. \end{cases}$$

We think of Y as a **censored** version of the Exponential random variable X: we cannot directly observe X, but we are able to gather some information about it (in this case, whether or not X is larger than 5.)

Observe that Y is a Bernoulli random variable. Thus, the statistical model for Y can be written $\left(\{0,1\}, \{\text{Ber}(f(\lambda))\}_{\lambda>0}\right)$ for some function f of λ. What is $f(\lambda)$?

(Type **lambda** for λ. Use the help button below for help with formula input).

$f(\lambda) =$ exp(-5*lambda)

✔ **Answer:** e^(-5*lambda)

$\exp(-5 \cdot \lambda)$

STANDARD NOTATION

Solution:

Note that $Y = 1$ if and only if $X > 5$. Hence, we need to compute the probability that $X > 5$. Recall that the density of $\mathrm{Exp}\,(\lambda)$ is given by $\lambda e^{-\lambda x}$. We just need to compute

$$P\,(X > 5) = \int_5^\infty \lambda e^{-\lambda x}\,dx = e^{-5\lambda}.$$

We conclude that if $X \sim \mathrm{Exp}\,(\lambda)$, then $Y \sim \mathrm{Ber}\,(e^{-5\lambda})$. Hence, $f\,(\lambda) = e^{-5\lambda}$.

Submit

You have used 1 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 2 Foundation of Inference:Lecture 3: Parametric Statistical Models / 8. Exercises on Statistical models

Add a Post

Show all posts ▼ by recent activity ▼

Notation of indicator function 6

Learn About Verified Certificates