Keener, Theoretical Statistics
Chapter 2: Exponential Families

Junhyeok Kil

July 7, 2025

2.4 Moments, Cumulants, and Generating Functions

Consider a random vector $T = (T_1, T_2, \ldots, T_s) \in \mathbb{R}^s$, which we can think of as the outcome of a random experiment. Let $u = (u_1, u_2, \ldots, u_s) \in \mathbb{R}^s$ be a constant vector.

2.4 Moments, Cumulants, and Generating Functions

Consider a random vector $T = (T_1, T_2, \dots, T_s) \in \mathbb{R}^s$, which we can think of as the outcome of a random experiment. Let $u = (u_1, u_2, \dots, u_s) \in \mathbb{R}^s$ be a constant vector.

Definition: Moment Generating Function (MGF)

The MGF of the random vector T is defined as:

$$M_T(u) = \mathbb{E}[e^{u_1 T_1 + \cdots + u_s T_s}] = \mathbb{E}[e^{u^T T}]$$

for $u \in \mathbb{R}^s$.

2.4 Moments, Cumulants, and Generating Functions

Consider a random vector $T = (T_1, T_2, \ldots, T_s) \in \mathbb{R}^s$, which we can think of as the outcome of a random experiment. Let $u = (u_1, u_2, \ldots, u_s) \in \mathbb{R}^s$ be a constant vector.

Definition: Moment Generating Function (MGF)

The MGF of the random vector T is defined as:

$$M_T(u) = \mathbb{E}[e^{u_1 T_1 + \cdots + u_s T_s}] = \mathbb{E}[e^{u^T T}]$$

for $u \in \mathbb{R}^s$.

Definition: Cumulant Generating Function (CGF)

The CGF of the random vector T is the natural logarithm of the MGF:

$$K_T(u) = \log M_T(u)$$

Uniqueness of MGFs and Definition of Moments

Lemma (Uniqueness of MGFs)

If the moment generating functions $M_X(u)$ and $M_Y(u)$ for two random vectors X and Y are finite and agree for u in some set with a non-empty interior, then X and Y have the same distribution, i.e., $p_X = p_Y$.

Uniqueness of MGFs and Definition of Moments

Lemma (Uniqueness of MGFs)

If the moment generating functions $M_X(u)$ and $M_Y(u)$ for two random vectors X and Y are finite and agree for u in some set with a non-empty interior, then X and Y have the same distribution, i.e., $p_X = p_Y$.

Definition: Moments

The expectations of products of powers of $T_1, \ldots, T_s$ are called the moments of T, denoted:

$$\alpha_{r_1,\ldots,r_s} = \mathbb{E}[T_1^{r_1} \times T_2^{r_2} \times \cdots \times T_s^{r_s}]$$

Uniqueness of MGFs and Definition of Moments

Lemma (Uniqueness of MGFs)

If the moment generating functions $M_X(u)$ and $M_Y(u)$ for two random vectors X and Y are finite and agree for u in some set with a non-empty interior, then X and Y have the same distribution, i.e., $p_X = p_Y$.

Definition: Moments

The expectations of products of powers of $T_1, \ldots, T_s$ are called the moments of T, denoted:

$$\alpha_{r_1,\ldots,r_s} = \mathbb{E}[T_1^{r_1} \times T_2^{r_2} \times \cdots \times T_s^{r_s}]$$

Example for $T = (T_1, T_2)$

Uniqueness of MGFs and Definition of Moments

Lemma (Uniqueness of MGFs)

If the moment generating functions $M_X(u)$ and $M_Y(u)$ for two random vectors X and Y are finite and agree for u in some set with a non-empty interior, then X and Y have the same distribution, i.e., $p_X = p_Y$.

Definition: Moments

The expectations of products of powers of $T_1, \ldots, T_s$ are called the moments of T, denoted:

$$\alpha_{r_1,\ldots,r_s} = \mathbb{E}[T_1^{r_1} \times T_2^{r_2} \times \cdots \times T_s^{r_s}]$$

Example for $T = (T_1, T_2)$

- $\alpha_{1,0} = \mathbb{E}[T_1^1 T_2^0] = \mathbb{E}[T_1]$ (Mean)

Uniqueness of MGFs and Definition of Moments

Lemma (Uniqueness of MGFs)

If the moment generating functions $M_X(u)$ and $M_Y(u)$ for two random vectors X and Y are finite and agree for u in some set with a non-empty interior, then X and Y have the same distribution, i.e., $p_X = p_Y$.

Definition: Moments

The expectations of products of powers of $T_1, \ldots, T_s$ are called the moments of T, denoted:

$$\alpha_{r_1,\ldots,r_s} = \mathbb{E}[T_1^{r_1} \times T_2^{r_2} \times \cdots \times T_s^{r_s}]$$

Example for $T = (T_1, T_2)$

- $\alpha_{1,0} = \mathbb{E}[T_1^1 T_2^0] = \mathbb{E}[T_1]$ (Mean)
- $\alpha_{0,2} = \mathbb{E}[T_1^0 T_2^2] = \mathbb{E}[T_2^2]$ (2nd moment of T_2)

Uniqueness of MGFs and Definition of Moments

Lemma (Uniqueness of MGFs)

If the moment generating functions $M_X(u)$ and $M_Y(u)$ for two random vectors X and Y are finite and agree for u in some set with a non-empty interior, then X and Y have the same distribution, i.e., $p_X = p_Y$.

Definition: Moments

The expectations of products of powers of $T_1, \ldots, T_s$ are called the moments of T, denoted:

$$\alpha_{r_1,\ldots,r_s} = \mathbb{E}[T_1^{r_1} \times T_2^{r_2} \times \cdots \times T_s^{r_s}]$$

Example for $T = (T_1, T_2)$

- $\alpha_{1,0} = \mathbb{E}[T_1^1 T_2^0] = \mathbb{E}[T_1]$ (Mean)
- $\alpha_{0,2} = \mathbb{E}[T_1^0 T_2^2] = \mathbb{E}[T_2^2]$ (2nd moment of T_2)
- $\alpha_{1,1} = \mathbb{E}[T_1^1 T_2^1] = \mathbb{E}[T_1 T_2]$ (Mixed moment)

Generating Moments and Cumulants

Theorem (Moments from MGF)

If the MGF $M_T(u)$ is finite in some neighborhood of the origin, then $M_T(u)$ has continuous derivatives of all orders at the origin. The moments of T can be found by differentiating $M_T(u)$ and evaluating at $u = 0$:

$$\alpha_{r_1,\ldots,r_s} = \frac{\partial^{r_1+\cdots+r_s}}{\partial u_1^{r_1} \ldots \partial u_s^{r_s}} M_T(u)\bigg|_{u=0}$$

Definition: Cumulants

The corresponding derivatives of the Cumulant Generating Function $K_T(u)$ are called the **cumulants**, denoted κ:

$$\kappa_{r_1,\ldots,r_s} = \frac{\partial^{r_1+\cdots+r_s}}{\partial u_1^{r_1} \ldots \partial u_s^{r_s}} K_T(u)\bigg|_{u=0}$$

Proof of the Moment Generation Theorem (1/3)

Given Moment Generating Function (MGF), $M_T(u) = \mathbb{E}[e^{u \cdot T}]$, and the Taylor series expansion for an exponential function, $e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$.
Expand the term $e^{u \cdot T}$:

$$
\begin{aligned}
e^{u \cdot T} &= e^{u_1 T_1 + \cdots + u_s T_s} \\
&= e^{u_1 T_1} \times e^{u_2 T_2} \times \cdots \times e^{u_s T_s} \\
&= \left(\sum_{r_1=0}^{\infty} \frac{(u_1 T_1)^{r_1}}{r_1!} \right) \times \cdots \times \left(\sum_{r_s=0}^{\infty} \frac{(u_s T_s)^{r_s}}{r_s!} \right) \\
&= \sum_{r_1, \ldots, r_s = 0}^{\infty} \frac{u_1^{r_1} \ldots u_s^{r_s}}{r_1! \ldots r_s!} T_1^{r_1} \ldots T_s^{r_s}
\end{aligned}
$$

Proof of the Moment Generation Theorem (2/3)

Take the expectation to find the MGF. By the linearity of expectation:

$$M_T(u) = \mathbb{E}[e^{u \cdot T}] = \mathbb{E}\left[\sum_{r_1,\ldots,r_s=0}^{\infty} \frac{u_1^{r_1} \ldots u_s^{r_s}}{r_1! \ldots r_s!} T_1^{r_1} \ldots T_s^{r_s}\right]$$

$$= \sum_{r_1,\ldots,r_s=0}^{\infty} \frac{u_1^{r_1} \ldots u_s^{r_s}}{r_1! \ldots r_s!} \mathbb{E}[T_1^{r_1} \ldots T_s^{r_s}]$$

Substitute the definition of moments, $\alpha_{r_1,\ldots,r_s} = \mathbb{E}[T_1^{r_1} \ldots T_s^{r_s}]$:

$$M_T(u) = \sum_{r_1,\ldots,r_s=0}^{\infty} \frac{\alpha_{r_1,\ldots,r_s}}{r_1! \ldots r_s!} u_1^{r_1} \ldots u_s^{r_s}$$

Proof of the Moment Generation Theorem (3/3)

We have established the power series representation of the MGF:

$$M_T(u) = \sum_{r_1,\dots,r_s=0}^{\infty} \frac{\alpha_{r_1,\dots,r_s}}{r_1!\dots r_s!} u_1^{r_1}\dots u_s^{r_s}$$

This is a multivariate Taylor series for $M_T(u)$ around the origin. The coefficients of this series are determined by the partial derivatives of the function. By differentiating this series term-by-term with respect to each u_i and then setting all $u_i = 0$, we isolate the coefficient corresponding to the desired moment:

$$\frac{\partial^{r_1+\dots+r_s}}{\partial u_1^{r_1}\dots\partial u_s^{r_s}} M_T(u)\bigg|_{u=0} = \alpha_{r_1,\dots,r_s}$$

Example: Cumulants of a Single Variable

Let T be a single random variable ($s = 1$), so u is a scalar. We have $K_T(u) = \log M_T(u)$.

First Cumulant (κ_1):

- Derivative: $K'_T(u) = \frac{M'_T(u)}{M_T(u)}$.
- At $u = 0$: $M_T(0) = \mathbb{E}[e^0] = 1$ and $M'_T(0) = \mathbb{E}[Te^0] = \mathbb{E}[T]$.
- Result: $\kappa_1 = K'_T(0) = \frac{\mathbb{E}[T]}{1} = \mathbb{E}[T]$.
- *The 1st cumulant is the mean.*

Second Cumulant (κ_2):

- Derivative: $K''_T(u) = \frac{M''_T(u)M_T(u) - (M'_T(u))^2}{M_T(u)^2}$.
- At $u = 0$: We also need $M''_T(0) = \mathbb{E}[T^2 e^0] = \mathbb{E}[T^2]$.
- Result: $\kappa_2 = K''_T(0) = \frac{\mathbb{E}[T^2] \cdot 1 - (\mathbb{E}[T])^2}{1^2} = \mathbb{E}[T^2] - (\mathbb{E}[T])^2$.
- *The 2nd cumulant is the variance, $Var(T)$.*

A Lemma on Independence

Generating functions are particularly useful in the study of sums of independent random variables. The following property is fundamental.

Lemma

Suppose X and Y are independent random variables. If X and Y are both positive, or if $\mathbb{E}[|X|]$ and $\mathbb{E}[|Y|]$ are both finite, then:

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Proof

View $|X||Y|$ as a function g of the joint random variable $Z = (X, Y)$, which has the product measure $P_X \times P_Y$ due to independence.

$$
\begin{aligned}
\mathbb{E}[|X||Y|] &= \int \int |x||y| \, d(P_X \times P_Y) \\
&= \int \left(\int |x||y| \, dP_X(x) \right) dP_Y(y) \quad \text{(by Fubini's Thm)} \\
&= \int |y| \left(\int |x| \, dP_X(x) \right) dP_Y(y) \\
&= \int |y| \mathbb{E}[|X|] \, dP_Y(y) \\
&= \mathbb{E}[|X|] \int |y| \, dP_Y(y) = \mathbb{E}[|X|]\mathbb{E}[|Y|]
\end{aligned}
$$

If $\mathbb{E}[|X|], \mathbb{E}[|Y|]$ are finite, then $\mathbb{E}[|X||Y|] < \infty$, then apply same steps omitting the absolute values.

Generating Functions of Sums

MGF and CGF of a Sum of Independent Vectors

The lemma on expectations extends by iteration to products of several independent variables. Suppose $T = Y_1 + \cdots + Y_n$, where $Y_1, \ldots, Y_n$ are independent random vectors.

MGF of a Sum

$$M_T(u) = \mathbb{E}[e^{u^T T}] = \mathbb{E}[e^{u^T(Y_1 + \cdots + Y_n)}]$$
$$= \mathbb{E}[e^{u^T Y_1} e^{u^T Y_2} \cdots e^{u^T Y_n}]$$

Since the Y_i vectors are independent, the random variables $e^{u^T Y_i}$ are also independent. Thus, the expectation of the product is the product of the expectations:

$$M_T(u) = \prod_{i=1}^{n} \mathbb{E}[e^{u^T Y_i}] = \prod_{i=1}^{n} M_{Y_i}(u)$$

Generating Functions of Sums

MGF and CGF of a Sum of Independent Vectors

Suppose $T = Y_1 + \cdots + Y_n$, where $Y_1, \ldots, Y_n$ are independent random vectors.

CGF of a Sum

Taking the logarithm gives a simpler relationship for the CGF:

$$K_T(u) = \log\left(\prod_{i=1}^{n} M_{Y_i}(u)\right) = \sum_{i=1}^{n} \log M_{Y_i}(u)$$

$$\implies K_T(u) = \sum_{i=1}^{n} K_{Y_i}(u)$$

The Additivity of Cumulants

We get cumulants by differentiating the CGF ($K_T(u)$) and evaluating at $u = 0$. Since $K_T(u)$ is a sum of individual CGFs, its derivatives will be the sum of the individual derivatives.

The Additivity Property

The cumulants of a sum of independent random vectors are the sums of the corresponding cumulants of the individual vectors.

The Additivity of Cumulants

We get cumulants by differentiating the CGF ($K_T(u)$) and evaluating at $u = 0$. Since $K_T(u)$ is a sum of individual CGFs, its derivatives will be the sum of the individual derivatives.

The Additivity Property

The cumulants of a sum of independent random vectors are the sums of the corresponding cumulants of the individual vectors.

Important Consequences

- **Mean (1st Cumulant):** The mean of a sum is the sum of the means.

$$\text{Mean}(T) = \sum_{i=1}^{n} \text{Mean}(Y_i)$$

The Additivity of Cumulants

We get cumulants by differentiating the CGF ($K_T(u)$) and evaluating at $u = 0$. Since $K_T(u)$ is a sum of individual CGFs, its derivatives will be the sum of the individual derivatives.

The Additivity Property

The cumulants of a sum of independent random vectors are the sums of the corresponding cumulants of the individual vectors.

Important Consequences

- **Mean (1st Cumulant):** The mean of a sum is the sum of the means.

$$\text{Mean}(T) = \sum_{i=1}^{n} \text{Mean}(Y_i)$$

- **Variance (2nd Cumulant):** The variance of a sum is the sum of the variances.

$$\text{Var}(T) = \sum_{i=1}^{n} \text{Var}(Y_i)$$

Generating Functions for Exponential Families

If X has a density from a canonical exponential family, we can find the MGF of its sufficient statistic $T = T(X)$ directly.

$$
\begin{aligned}
M_T(u) = \mathbb{E}_\eta[e^{u^T T}] &= \int e^{u^T T(x)} p_\eta(x)\, d\mu(x) \\
&= \int e^{u^T T(x)} \exp(\eta^T T(x) - A(\eta)) h(x)\, d\mu(x) \\
&= \int \exp((u+\eta)^T T(x) - A(\eta)) h(x)\, d\mu(x) \\
&= e^{-A(\eta)} \int \exp((u+\eta)^T T(x)) h(x)\, d\mu(x) \\
&= e^{-A(\eta)} \cdot e^{A(u+\eta)} \qquad \text{(by definition of } A(\cdot))
\end{aligned}
$$

This holds provided $(u+\eta)$ is in the natural parameter space Ξ.

The CGF for an Exponential Family

The Moment Generating Function is: $M_T(u) = e^{A(u+\eta)-A(\eta)}$.
Therefore, the Cumulant Generating Function is simply:

$$K_T(u) = A(u+\eta) - A(\eta)$$

The Power of the Log-Partition Function

The cumulants of T are the derivatives of $K_T(u)$ with respect to u, evaluated at $u = 0$.

$$\frac{\partial}{\partial u_j} K_T(u) = \frac{\partial}{\partial u_j} \left(A(u+\eta) - A(\eta) \right) = \left. \frac{\partial A(v)}{\partial v_j} \right|_{v=u+\eta}$$

Evaluating at $u = 0$ means we are simply taking the partial derivative of $A(\eta)$ with respect to η_j. This generalizes to all higher-order derivatives.

Main Result

The cumulants of the sufficient statistic T are the partial derivatives of the log-partition function $A(\eta)$ with respect to the natural parameters η.

$$\kappa_{r_1,\ldots,r_s}(T) = \frac{\partial^{r_1+\cdots+r_s}}{\partial \eta_1^{r_1} \ldots \partial \eta_s^{r_s}} A(\eta)$$

The Power of the Log-Partition Function

Moments from $A(\eta)$

The Power of the Log-Partition Function

Moments from $A(\eta)$

- **Mean:** $\mathbb{E}_\eta[T_i] = \dfrac{\partial A(\eta)}{\partial \eta_i}$

The Power of the Log-Partition Function

Moments from $A(\eta)$

- **Mean:** $\mathbb{E}_\eta[T_i] = \dfrac{\partial A(\eta)}{\partial \eta_i}$
- **Variance:** $\mathrm{Var}_\eta(T_i) = \dfrac{\partial^2 A(\eta)}{\partial \eta_i^2}$

The Power of the Log-Partition Function

Moments from $A(\eta)$

- **Mean:** $\mathbb{E}_\eta[T_i] = \dfrac{\partial A(\eta)}{\partial \eta_i}$
- **Variance:** $\mathrm{Var}_\eta(T_i) = \dfrac{\partial^2 A(\eta)}{\partial \eta_i^2}$
- **Covariance:** $\mathrm{Cov}_\eta(T_i, T_j) = \dfrac{\partial^2 A(\eta)}{\partial \eta_i \partial \eta_j}$

Example: Cumulants of the Poisson Distribution

If X has the Poisson distribution with mean λ, its probability mass function is:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{1}{x!} \exp(x \log \lambda - \lambda)$$

Example: Cumulants of the Poisson Distribution

If X has the Poisson distribution with mean λ, its probability mass function is:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{1}{x!} \exp(x \log \lambda - \lambda)$$

This is an exponential family. To find the canonical form, we set

the canonical parameter $\eta = \log \lambda$, which means $\lambda = e^\eta$.

$$P(X = x) = \frac{1}{x!} \exp(\eta x - e^\eta)$$

From this form, we identify:

Example: Cumulants of the Poisson Distribution

If X has the Poisson distribution with mean λ, its probability mass function is:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{1}{x!} \exp(x \log \lambda - \lambda)$$

This is an exponential family. To find the canonical form, we set

the canonical parameter $\eta = \log \lambda$, which means $\lambda = e^\eta$.

$$P(X = x) = \frac{1}{x!} \exp(\eta x - e^\eta)$$

From this form, we identify:

- Sufficient Statistic: $T(x) = x$.

Example: Cumulants of the Poisson Distribution

If X has the Poisson distribution with mean λ, its probability mass function is:

$$P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{1}{x!}\exp(x\log\lambda - \lambda)$$

This is an exponential family. To find the canonical form, we set

the canonical parameter $\eta = \log\lambda$, which means $\lambda = e^\eta$.

$$P(X=x) = \frac{1}{x!}\exp(\eta x - e^\eta)$$

From this form, we identify:

- Sufficient Statistic: $T(x) = x$.
- Log-Partition Function: $A(\eta) = e^\eta$.

Example: Cumulants of the Poisson Distribution

If X has the Poisson distribution with mean λ, its probability mass function is:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{1}{x!} \exp(x \log \lambda - \lambda)$$

This is an exponential family. To find the canonical form, we set

the canonical parameter $\eta = \log \lambda$, which means $\lambda = e^{\eta}$.

$$P(X = x) = \frac{1}{x!} \exp(\eta x - e^{\eta})$$

From this form, we identify:

- Sufficient Statistic: $T(x) = x$.
- Log-Partition Function: $A(\eta) = e^{\eta}$.

The cumulants of X are the derivatives of $A(\eta)$ with respect to η:

$$\kappa_k = \frac{d^k A(\eta)}{d\eta^k} = \frac{d^k e^{\eta}}{d\eta^k} = e^{\eta} = \lambda$$

Example: Normal Distribution with Fixed Variance

Consider the class of normal densities $N(\mu, \sigma^2)$ where μ varies but σ^2 is fixed. We can arrange the PDF as:

$$p_\mu(x) = \exp\left(\frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2}\right) \cdot \frac{\exp(-x^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}}$$

This is a canonical exponential family with:

Example: Normal Distribution with Fixed Variance

Consider the class of normal densities $N(\mu, \sigma^2)$ where μ varies but σ^2 is fixed. We can arrange the PDF as:

$$p_\mu(x) = \exp\left(\frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2}\right) \cdot \frac{\exp(-x^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}}$$

This is a canonical exponential family with:

- Sufficient Statistic: $T(x) = x$.

Example: Normal Distribution with Fixed Variance

Consider the class of normal densities $N(\mu, \sigma^2)$ where μ varies but σ^2 is fixed. We can arrange the PDF as:

$$p_\mu(x) = \exp\left(\frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2}\right) \cdot \frac{\exp(-x^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}}$$

This is a canonical exponential family with:

- Sufficient Statistic: $T(x) = x$.
- Canonical Parameter: $\eta = \mu/\sigma^2$.

Example: Normal Distribution with Fixed Variance

Consider the class of normal densities $N(\mu, \sigma^2)$ where μ varies but σ^2 is fixed. We can arrange the PDF as:

$$p_\mu(x) = \exp\left(\frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2}\right) \cdot \frac{\exp(-x^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}}$$

This is a canonical exponential family with:

- Sufficient Statistic: $T(x) = x$.
- Canonical Parameter: $\eta = \mu/\sigma^2$.
- Log-Partition Function: $A(\eta) = \frac{\mu^2}{2\sigma^2} = \frac{(\eta\sigma^2)^2}{2\sigma^2} = \frac{\sigma^2\eta^2}{2}$.

Example: Normal Distribution with Fixed Variance

Consider the class of normal densities $N(\mu, \sigma^2)$ where μ varies but σ^2 is fixed. We can arrange the PDF as:

$$p_\mu(x) = \exp\left(\frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2}\right) \cdot \frac{\exp(-x^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}}$$

This is a canonical exponential family with:

- Sufficient Statistic: $T(x) = x$.
- Canonical Parameter: $\eta = \mu/\sigma^2$.
- Log-Partition Function: $A(\eta) = \frac{\mu^2}{2\sigma^2} = \frac{(\eta\sigma^2)^2}{2\sigma^2} = \dfrac{\sigma^2\eta^2}{2}$.

The first two cumulants are derivatives of the quadratic function $A(\eta)$:

Example: Normal Distribution with Fixed Variance

Consider the class of normal densities $N(\mu, \sigma^2)$ where μ varies but σ^2 is fixed. We can arrange the PDF as:

$$p_\mu(x) = \exp\left(\frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2}\right) \cdot \frac{\exp(-x^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}}$$

This is a canonical exponential family with:

- Sufficient Statistic: $T(x) = x$.
- Canonical Parameter: $\eta = \mu/\sigma^2$.
- Log-Partition Function: $A(\eta) = \frac{\mu^2}{2\sigma^2} = \frac{(\eta\sigma^2)^2}{2\sigma^2} = \dfrac{\sigma^2\eta^2}{2}$.

The first two cumulants are derivatives of the quadratic function $A(\eta)$:

- $\kappa_1 = A'(\eta) = \sigma^2\eta = \sigma^2(\mu/\sigma^2) = \mu \quad$ (the mean).

Example: Normal Distribution with Fixed Variance

Consider the class of normal densities $N(\mu, \sigma^2)$ where μ varies but σ^2 is fixed. We can arrange the PDF as:

$$p_\mu(x) = \exp\left(\frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2}\right) \cdot \frac{\exp(-x^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}}$$

This is a canonical exponential family with:

- Sufficient Statistic: $T(x) = x$.
- Canonical Parameter: $\eta = \mu/\sigma^2$.
- Log-Partition Function: $A(\eta) = \frac{\mu^2}{2\sigma^2} = \frac{(\eta\sigma^2)^2}{2\sigma^2} = \dfrac{\sigma^2\eta^2}{2}$.

The first two cumulants are derivatives of the quadratic function $A(\eta)$:

- $\kappa_1 = A'(\eta) = \sigma^2\eta = \sigma^2(\mu/\sigma^2) = \mu \quad$ (the mean).
- $\kappa_2 = A''(\eta) = \sigma^2 \quad$ (the variance).

Example: Normal Distribution with Fixed Variance

Consider the class of normal densities $N(\mu, \sigma^2)$ where μ varies but σ^2 is fixed. We can arrange the PDF as:

$$p_\mu(x) = \exp\left(\frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2}\right) \cdot \frac{\exp(-x^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}}$$

This is a canonical exponential family with:

- Sufficient Statistic: $T(x) = x$.
- Canonical Parameter: $\eta = \mu/\sigma^2$.
- Log-Partition Function: $A(\eta) = \frac{\mu^2}{2\sigma^2} = \frac{(\eta\sigma^2)^2}{2\sigma^2} = \dfrac{\sigma^2\eta^2}{2}$.

The first two cumulants are derivatives of the quadratic function $A(\eta)$:

- $\kappa_1 = A'(\eta) = \sigma^2\eta = \sigma^2(\mu/\sigma^2) = \mu \quad$ (the mean).
- $\kappa_2 = A''(\eta) = \sigma^2 \quad$ (the variance).

Since $A(\eta)$ is quadratic, all higher-order cumulants $(\kappa_3, \kappa_4, \dots)$ are zero for the Normal distribution.

From Cumulants to Moments

To calculate moments from cumulants for a single random variable ($s = 1$), we can repeatedly differentiate the identity $M_T(u) = e^{K_T(u)}$ and evaluate the results at $u = 0$.

Recalling that at $u = 0$, $\mathbb{E}[T^n] = M_T^{(n)}(0)$ and $\kappa_n = K_T^{(n)}(0)$, this gives the following relationships:

Moments as Functions of Cumulants

$$
\begin{aligned}
\mathbb{E}[T] &= \kappa_1 \\
\mathbb{E}[T^2] &= \kappa_2 + \kappa_1^2 \\
\mathbb{E}[T^3] &= \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3 \\
\mathbb{E}[T^4] &= \kappa_4 + 3\kappa_2^2 + 4\kappa_1\kappa_3 + 6\kappa_1^2\kappa_2 + \kappa_1^4
\end{aligned}
$$

These expressions can also be solved to express cumulants as functions of moments.

Examples: Higher Moments from Cumulants

Poisson(λ)

Recall that for a Poisson distribution, all cumulants are $\kappa_n = \lambda$.

Examples: Higher Moments from Cumulants

Poisson(λ)

Recall that for a Poisson distribution, all cumulants are $\kappa_n = \lambda$.

- $\mathbb{E}[X^2] = \kappa_2 + \kappa_1^2 = \lambda + \lambda^2$

Examples: Higher Moments from Cumulants

Poisson(λ)

Recall that for a Poisson distribution, all cumulants are $\kappa_n = \lambda$.

- $\mathbb{E}[X^2] = \kappa_2 + \kappa_1^2 = \lambda + \lambda^2$
- $\mathbb{E}[X^3] = \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3 = \lambda + 3\lambda^2 + \lambda^3$

Examples: Higher Moments from Cumulants

Poisson(λ)

Recall that for a Poisson distribution, all cumulants are $\kappa_n = \lambda$.

- $\mathbb{E}[X^2] = \kappa_2 + \kappa_1^2 = \lambda + \lambda^2$
- $\mathbb{E}[X^3] = \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3 = \lambda + 3\lambda^2 + \lambda^3$
- $\mathbb{E}[X^4] = \kappa_4 + 3\kappa_2^2 + 4\kappa_1\kappa_3 + \cdots =$
  $\lambda + 3\lambda^2 + 4\lambda^2 + 6\lambda^3 + \lambda^4 = \lambda + 7\lambda^2 + 6\lambda^3 + \lambda^4$

Normal(μ, σ^2)

Recall that for a Normal distribution, $\kappa_1 = \mu$, $\kappa_2 = \sigma^2$, and $\kappa_{n\geq 3} = 0$.

Examples: Higher Moments from Cumulants

Poisson(λ)

Recall that for a Poisson distribution, all cumulants are $\kappa_n = \lambda$.

- $\mathbb{E}[X^2] = \kappa_2 + \kappa_1^2 = \lambda + \lambda^2$
- $\mathbb{E}[X^3] = \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3 = \lambda + 3\lambda^2 + \lambda^3$
- $\mathbb{E}[X^4] = \kappa_4 + 3\kappa_2^2 + 4\kappa_1\kappa_3 + \cdots =$
  $\lambda + 3\lambda^2 + 4\lambda^2 + 6\lambda^3 + \lambda^4 = \lambda + 7\lambda^2 + 6\lambda^3 + \lambda^4$

Normal(μ, σ^2)

Recall that for a Normal distribution, $\kappa_1 = \mu$, $\kappa_2 = \sigma^2$, and $\kappa_{n\geq 3} = 0$.

- $\mathbb{E}[X^3] = \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3 = 0 + 3\mu\sigma^2 + \mu^3$

Examples: Higher Moments from Cumulants

Poisson(λ)

Recall that for a Poisson distribution, all cumulants are $\kappa_n = \lambda$.

- $\mathbb{E}[X^2] = \kappa_2 + \kappa_1^2 = \lambda + \lambda^2$
- $\mathbb{E}[X^3] = \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3 = \lambda + 3\lambda^2 + \lambda^3$
- $\mathbb{E}[X^4] = \kappa_4 + 3\kappa_2^2 + 4\kappa_1\kappa_3 + \cdots =$
  $\lambda + 3\lambda^2 + 4\lambda^2 + 6\lambda^3 + \lambda^4 = \lambda + 7\lambda^2 + 6\lambda^3 + \lambda^4$

Normal(μ, σ^2)

Recall that for a Normal distribution, $\kappa_1 = \mu$, $\kappa_2 = \sigma^2$, and $\kappa_{n\geq 3} = 0$.

- $\mathbb{E}[X^3] = \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3 = 0 + 3\mu\sigma^2 + \mu^3$
- $\mathbb{E}[X^4] = \kappa_4 + 3\kappa_2^2 + 4\kappa_1\kappa_3 + 6\kappa_1^2\kappa_2 + \kappa_1^4 =$
  $0 + 3(\sigma^2)^2 + 0 + 6\mu^2\sigma^2 + \mu^4 = 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4$

CGF of a Shifted Random Vector

Consider shifting a random vector T by a constant vector $c \in \mathbb{R}^s$. The MGF of the new vector $T + c$ is:

$$
\begin{aligned}
M_{T+c}(u) &= \mathbb{E}[e^{u^T(T+c)}] \\
&= \mathbb{E}[e^{u^T T + u^T c}] \\
&= e^{u^T c} \cdot \mathbb{E}[e^{u^T T}] \\
&= e^{u^T c} M_T(u)
\end{aligned}
$$

CGF of a Shifted Random Vector

CGF Shift Property

Taking the logarithm reveals a simple additive relationship for the CGF:

$$K_{T+c}(u) = \log(e^{u^T c} M_T(u)) = u^T c + K_T(u)$$

This shows that shifting the random variable only adds a linear term to the CGF, which only affects the first cumulant (the mean) and leaves all higher-order cumulants (variance, skewness, etc.) unchanged.

Cumulants and Central Moments

The shift property $K_{T+c}(u) = u^T c + K_T(u)$ implies that for any derivative of order $j \geq 2$, the linear term $u^T c$ vanishes. This means that cumulants of order 2 and higher are **shift-invariant**. Specifically, they are invariant to centering the random variable:

$$\kappa_j(T) = \kappa_j(T - \mathbb{E}[T]) \quad \text{for } j \geq 2$$

This invariance provides a direct link between cumulants and

central moments.

Cumulants and Central Moments

Relationship to Central Moments

For a centered random variable $Y = T - \mathbb{E}[T]$, the first cumulant $\kappa_1(Y)$ is 0. This simplifies the moment-cumulant formulas:

- **Third Cumulant:** $\kappa_3 = \mathbb{E}[(T - \mathbb{E}[T])^3]$. The 3rd cumulant measures skewness.
- **Fourth Cumulant:** $\mathbb{E}[(T - \mathbb{E}[T])^4] = \kappa_4 + 3\kappa_2^2$. This is often rearranged to define the 4th cumulant:

$$\kappa_4 = \mathbb{E}[(T - \mathbb{E}[T])^4] - 3(\mathrm{Var}(T))^2$$

  This is also known as the *excess kurtosis*.

Higher Dimensions

In the multi-dimensional case, the low-order cumulants correspond to the means and the covariance matrix of the vector T.

Exercise 2.1

Consider independent Bernoulli trials with a success probability of p. Let X be the number of failures before the first success. The probability mass function (PMF), using θ for the success probability, is:

$$P(X = x|\theta) = \theta(1-\theta)^x, \quad x = 0, 1, 2, \ldots$$

Goals:

a) Show the geometric distribution is an exponential family.
b) Write its density in canonical form and identify its components.
c) Find the mean of the distribution using differentiation.
d) Analyze the joint distribution of i.i.d. samples.

a) An Exponential Family Member

A distribution belongs to the exponential family if its PMF can be written as:

$$p(x|\theta) = h(x)\exp(\eta(\theta)T(x) - B(\theta))$$

a) An Exponential Family Member

A distribution belongs to the exponential family if its PMF can be written as:

$$p(x|\theta) = h(x)\exp(\eta(\theta)T(x) - B(\theta))$$

For the geometric PMF, $p(x|\theta) = \theta(1-\theta)^x$:

$$p(x|\theta) = \exp\left(\log(\theta(1-\theta)^x)\right) \quad = \exp\left(\log(\theta) + x\log(1-\theta)\right)$$

a) An Exponential Family Member

A distribution belongs to the exponential family if its PMF can be written as:

$$p(x|\theta) = h(x) \exp(\eta(\theta) T(x) - B(\theta))$$

For the geometric PMF, $p(x|\theta) = \theta(1-\theta)^x$:

$$p(x|\theta) = \exp\left(\log(\theta(1-\theta)^x)\right) \quad = \exp\left(\log(\theta) + x\log(1-\theta)\right)$$

This form matches, with the sufficient statistic being $T(x) = x$ and $h(x) = 1$.

b) The Canonical Form

The canonical form is $p(x|\eta) = h(x)\exp(\eta T(x) - A(\eta))$. We need to express our PMF using the natural parameter η.

Canonical Parameter η

We have $\eta = \log(1 - \theta)$.

b) The Canonical Form

The canonical form is $p(x|\eta) = h(x)\exp(\eta T(x) - A(\eta))$. We need to express our PMF using the natural parameter η.

Canonical Parameter η

We have $\eta = \log(1-\theta)$. Solving for θ gives:

$$e^\eta = 1 - \theta \implies \theta = 1 - e^\eta$$

Log-Partition Function $A(\eta)$

Substituting θ back into the log-PMF from the previous slide:

$$\log(1-e^\eta) + x\eta = \eta x - (-\log(1-e^\eta))$$

b) The Canonical Form

The canonical form is $p(x|\eta) = h(x)\exp(\eta T(x) - A(\eta))$. We need to express our PMF using the natural parameter η.

Canonical Parameter η

We have $\eta = \log(1-\theta)$. Solving for θ gives:

$$e^\eta = 1 - \theta \implies \theta = 1 - e^\eta$$

Log-Partition Function $A(\eta)$

Substituting θ back into the log-PMF from the previous slide:

$$\log(1 - e^\eta) + x\eta = \eta x - (-\log(1 - e^\eta))$$

Thus, the log-partition function is:

$$A(\eta) = -\log(1 - e^\eta)$$

c) Finding the Mean via Differentiation

For an exponential family, the mean of the sufficient statistic is the first derivative of the log-partition function, $A(\eta)$.

$$E[X] = A'(\eta)$$

c) Finding the Mean via Differentiation

For an exponential family, the mean of the sufficient statistic is the first derivative of the log-partition function, $A(\eta)$.

$$E[X] = A'(\eta)$$

$$\begin{aligned} A'(\eta) &= \frac{d}{d\eta}\left(-\log(1-e^{\eta})\right) \\ &= -\frac{1}{1-e^{\eta}} \cdot (-e^{\eta}) = \frac{e^{\eta}}{1-e^{\eta}} \end{aligned}$$

c) Finding the Mean via Differentiation

For an exponential family, the mean of the sufficient statistic is the first derivative of the log-partition function, $A(\eta)$.

$$E[X] = A'(\eta)$$

$$\begin{aligned} A'(\eta) &= \frac{d}{d\eta}\left(-\log(1-e^{\eta})\right) \\ &= -\frac{1}{1-e^{\eta}} \cdot (-e^{\eta}) = \frac{e^{\eta}}{1-e^{\eta}} \end{aligned}$$

Converting back to θ (since $e^{\eta} = 1-\theta$):

$$E[X] = \frac{1-\theta}{\theta}$$

c) Finding the Variance via Differentiation

Similarly, the variance is the second derivative of $A(\eta)$.

$$Var(X) = A''(\eta)$$

$$\begin{aligned}
A''(\eta) &= \frac{d}{d\eta}\left(\frac{e^\eta}{1-e^\eta}\right) \\
&= \frac{e^\eta(1-e^\eta) - e^\eta(-e^\eta)}{(1-e^\eta)^2} \\
&= \frac{e^\eta}{(1-e^\eta)^2}
\end{aligned}$$

c) Finding the Variance via Differentiation

Similarly, the variance is the second derivative of $A(\eta)$.

$$Var(X) = A''(\eta)$$

$$\begin{aligned} A''(\eta) &= \frac{d}{d\eta}\left(\frac{e^\eta}{1-e^\eta}\right) \\ &= \frac{e^\eta(1-e^\eta) - e^\eta(-e^\eta)}{(1-e^\eta)^2} \\ &= \frac{e^\eta}{(1-e^\eta)^2} \end{aligned}$$

Converting back to θ:

$$Var(X) = \frac{1-\theta}{\theta^2}$$

d) Joint Distribution of i.i.d. Samples

For $X_1, \ldots, X_n$ i.i.d. from Geometric(θ), the joint PMF is:

$$\begin{aligned} p(\mathbf{x}|\theta) &= \prod_{i=1}^{n} \theta(1-\theta)^{x_i} \\ &= \theta^n (1-\theta)^{\sum x_i} \\ &= \exp\left(n\log(\theta) + \left(\sum x_i\right)\log(1-\theta)\right) \end{aligned}$$

d) Joint Distribution of i.i.d. Samples

For $X_1, \ldots, X_n$ i.i.d. from Geometric(θ), the joint PMF is:

$$
\begin{aligned}
p(\mathbf{x}|\theta) &= \prod_{i=1}^{n} \theta(1-\theta)^{x_i} \\
&= \theta^n (1-\theta)^{\sum x_i} \\
&= \exp\left(n \log(\theta) + \left(\sum x_i \right) \log(1-\theta) \right)
\end{aligned}
$$

In terms of $\eta = \log(1-\theta)$, this becomes:

$$
p(\mathbf{x}|\eta) = \exp\left(\eta \left(\sum x_i \right) + n \log(1 - e^{\eta}) \right)
$$

d) Joint Distribution of i.i.d. Samples

For $X_1, \ldots, X_n$ i.i.d. from Geometric(θ), the joint PMF is:

$$
\begin{aligned}
p(\mathbf{x}|\theta) &= \prod_{i=1}^{n} \theta(1-\theta)^{x_i} \\
&= \theta^n (1-\theta)^{\sum x_i} \\
&= \exp\left(n \log(\theta) + \left(\sum x_i\right) \log(1-\theta)\right)
\end{aligned}
$$

In terms of $\eta = \log(1-\theta)$, this becomes:

$$
p(\mathbf{x}|\eta) = \exp\left(\eta \left(\sum x_i\right) + n \log(1 - e^{\eta})\right)
$$

This is an exponential family with sufficient statistic $T(\mathbf{x}) = \sum_{i=1}^{n} X_i$ and log-partition function $A_n(\eta) = -n \log(1 - e^{\eta}) = nA(\eta)$.

d) Mean and Variance of the Sum

Using the new log-partition function $A_n(\eta) = nA(\eta)$, we can find the moments of the sufficient statistic $T = \sum X_i$.

Mean of T

$$\begin{aligned} E[T] &= \frac{dA_n(\eta)}{d\eta} = nA'(\eta) \\ &= n\frac{1-\theta}{\theta} \end{aligned}$$

d) Mean and Variance of the Sum

Using the new log-partition function $A_n(\eta) = nA(\eta)$, we can find the moments of the sufficient statistic $T = \sum X_i$.

Mean of T

$$E[T] = \frac{dA_n(\eta)}{d\eta} = nA'(\eta)$$
$$= n\frac{1-\theta}{\theta}$$

Variance of T

$$Var(T) = \frac{d^2 A_n(\eta)}{d\eta^2} = nA''(\eta)$$
$$= n\frac{1-\theta}{\theta^2}$$

Exercise 2.23

Suppose $Z \sim N(0,1)$. Find the first four cumulants of the random variable $X = Z^2$.

Hint

Consider the exponential family form of the more general distribution $N(0, \sigma^2)$.

The Exponential Family Formulation

The PDF of $Y \sim N(0, \sigma^2)$ is:

$$p(y; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2}{2\sigma^2}\right)$$

The Exponential Family Formulation

The PDF of $Y \sim N(0, \sigma^2)$ is:

$$p(y; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2}{2\sigma^2}\right)$$

We can rewrite this to identify its exponential family components:

$$p(y; \sigma^2) = \frac{1}{\sqrt{2\pi}} \exp\left(\left(\frac{-1}{2\sigma^2}\right) y^2 - \frac{1}{2}\log(\sigma^2)\right)$$

The Exponential Family Formulation

The PDF of $Y \sim N(0, \sigma^2)$ is:

$$p(y; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2}{2\sigma^2}\right)$$

We can rewrite this to identify its exponential family components:

$$p(y; \sigma^2) = \frac{1}{\sqrt{2\pi}} \exp\left(\left(\frac{-1}{2\sigma^2}\right) y^2 - \frac{1}{2}\log(\sigma^2)\right)$$

- Sufficient Statistic: $T(y) = y^2$
- Natural Parameter: $\eta = -\frac{1}{2\sigma^2}$

The cumulant-generating function (CGF) of the sufficient statistic $T(y) = y^2$ can be derived from this form.

Deriving the CGF for $T = Y^2$

The exponential family form is:

$$p(y) = h(y)\exp\left(\eta y^2 - A(\eta)\right), \quad h(y) = \frac{1}{\sqrt{2\pi}}, \quad A(\eta) = \frac{1}{2}\log\left(-\frac{1}{\eta}\right)$$

Define the moment-generating function (MGF) of $T = Y^2$:

$$M_T(u) = \mathbb{E}[e^{uY^2}] = \exp(-A(\eta))\int h(y)\exp\left((\eta+u)y^2\right)dy$$

We know that

$$M_T(u) = \exp\left(A(\eta+u) - A(\eta)\right)$$

Compute the CGF as the log of the MGF:

$$K_T(u) = A(\eta+u) - A(\eta) = \frac{1}{2}\log\left(\frac{-1/\eta}{-1/(\eta+u)}\right) = -\frac{1}{2}\log\left(1+\frac{u}{\eta}\right)$$

The Cumulant-Generating Function (CGF)

The CGF of the sufficient statistic $T = Y^2$ for $Y \sim N(0, \sigma^2)$ is found to be:

$$K_T(u) = -\frac{1}{2} \log\left(1 + \frac{u}{\eta}\right)$$

The Cumulant-Generating Function (CGF)

The CGF of the sufficient statistic $T = Y^2$ for $Y \sim N(0, \sigma^2)$ is found to be:

$$K_T(u) = -\frac{1}{2} \log \left(1 + \frac{u}{\eta} \right)$$

We are interested in the case where $Z \sim N(0, 1)$, so we set $\sigma^2 = 1$. This gives a natural parameter of:

$$\eta = -\frac{1}{2(1)^2} = -\frac{1}{2}$$

The Cumulant-Generating Function (CGF)

The CGF of the sufficient statistic $T = Y^2$ for $Y \sim N(0, \sigma^2)$ is found to be:

$$K_T(u) = -\frac{1}{2} \log \left(1 + \frac{u}{\eta} \right)$$

We are interested in the case where $Z \sim N(0, 1)$, so we set $\sigma^2 = 1$. This gives a natural parameter of:

$$\eta = -\frac{1}{2(1)^2} = -\frac{1}{2}$$

Substituting $\eta = -1/2$ into the CGF gives the CGF for Z^2:

$$\begin{aligned} K_{Z^2}(u) &= -\frac{1}{2} \log \left(1 + \frac{u}{-1/2} \right) \\ &= -\frac{1}{2} \log(1 - 2u) \end{aligned}$$

Calculating the Cumulants

The k-th cumulant, κ_k, is the k-th derivative of the CGF, evaluated at $u = 0$.

$$K(u) = -\frac{1}{2}\log(1 - 2u)$$

Calculating the Cumulants

The k-th cumulant, κ_k, is the k-th derivative of the CGF, evaluated at $u = 0$.

$$K(u) = -\frac{1}{2}\log(1-2u)$$

First Derivative (κ_1)

$K'(u) = -\frac{1}{2} \cdot \frac{-2}{1-2u} = (1-2u)^{-1}$
$\kappa_1 = K'(0) = 1$

Calculating the Cumulants

The k-th cumulant, κ_k, is the k-th derivative of the CGF, evaluated at $u = 0$.

$$K(u) = -\frac{1}{2}\log(1 - 2u)$$

First Derivative (κ_1)

$K'(u) = -\frac{1}{2} \cdot \frac{-2}{1-2u} = (1 - 2u)^{-1}$
$\kappa_1 = K'(0) = 1$

Second Derivative (κ_2)

$K''(u) = (-1)(1 - 2u)^{-2}(-2) = 2(1 - 2u)^{-2}$
$\kappa_2 = K''(0) = 2$

Calculating the Cumulants (cont.)

Continuing with the derivatives of $K(u) = -\frac{1}{2}\log(1-2u)$.

Third Derivative (κ_3)

$K'''(u) = 2(-2)(1-2u)^{-3}(-2) = 8(1-2u)^{-3}$
$\kappa_3 = K'''(0) = 8$

Calculating the Cumulants (cont.)

Continuing with the derivatives of $K(u) = -\frac{1}{2}\log(1-2u)$.

Third Derivative (κ_3)

$K'''(u) = 2(-2)(1-2u)^{-3}(-2) = 8(1-2u)^{-3}$
$\kappa_3 = K'''(0) = 8$

Fourth Derivative (κ_4)

$K^{(4)}(u) = 8(-3)(1-2u)^{-4}(-2) = 48(1-2u)^{-4}$
$\kappa_4 = K^{(4)}(0) = 48$

Summary of Results

For the random variable $X = Z^2$ where $Z \sim N(0,1)$, the first four cumulants are:

- $\kappa_1 = E[X] = 1$
- $\kappa_2 = \mathrm{Var}(X) = 2$
- $\kappa_3 = 8$
- $\kappa_4 = 48$

3.1 Models, Estimators, and Risk Functions

In inferential statistics, we aim to learn about an unknown parameter θ from observed data X.

The **parameter** θ and **data** X are related through a statistical **model**. When the parameter's value is θ, the data's distribution is denoted by P_θ.

Formal Definition

A model $\mathcal{P}$ is a set of distributions for the data X, indexed by the parameter θ:

$$\mathcal{P} = \{P_\theta : \theta \in \Omega\}$$

where Ω is the **parameter space**, representing all possible values for θ.

Example: A Coin Toss Experiment

Suppose we model a coin toss that is performed 100 times, with independent trials and a common probability θ of landing heads.

- **Data:** The total number of heads, X.
- **Distribution:** X follows a Binomial distribution, $X \sim \text{Binom}(100, \theta)$.
- **Parameter Space:** $\theta \in [0, 1]$.

The statistical model $\mathcal{P}$ is the set of all binomial distributions with 100 trials:

$$\mathcal{P} = \{P_\theta = \text{Binom}(100, \theta) \mid \theta \in [0, 1]\}$$

Estimation: Guessing the Parameter

A primary goal of statistical inference is **estimation**. We want to find a function of the data, called a **statistic**, that gives a good guess for the true value of θ.

Estimator

An **estimator** $\delta(X)$ is a function of the data used to estimate a parameter θ (or a function of it, $g(\theta)$). The resulting value, $\delta(X)$, is the **estimate**.

Estimation: Guessing the Parameter

A primary goal of statistical inference is **estimation**. We want to find a function of the data, called a **statistic**, that gives a good guess for the true value of θ.

Estimator

An **estimator** $\delta(X)$ is a function of the data used to estimate a parameter θ (or a function of it, $g(\theta)$). The resulting value, $\delta(X)$, is the **estimate**.

Example continued: For the coin toss experiment, a natural estimator for the probability of heads θ is the sample proportion:

$$\delta(X) = \frac{X}{100}$$

How Good Is an Estimate?

To judge the quality of an estimator, we need to quantify the "cost" or "error" of a bad estimate. This is done using a **loss function**.

Loss Function $L(\theta, d)$

A loss function $L(\theta, d)$ measures the penalty associated with estimating the true parameter value θ with an estimate d.

- $L(\theta, \theta) = 0$ (no loss for a perfect estimate)
- $L(\theta, d) \geq 0$ for all $d \neq \theta$

How Good Is an Estimate?

To judge the quality of an estimator, we need to quantify the "cost" or "error" of a bad estimate. This is done using a **loss function**.

Loss Function $L(\theta, d)$

A loss function $L(\theta, d)$ measures the penalty associated with estimating the true parameter value θ with an estimate d.

- $L(\theta, \theta) = 0$ (no loss for a perfect estimate)
- $L(\theta, d) \geq 0$ for all $d \neq \theta$

A very common choice is the **squared error loss**:

$$L(\theta, d) = (\theta - d)^2$$

Average Loss: The Risk Function

The loss $L(\theta, \delta(X))$ is random because the data X is random. We can be unlucky and get a large loss even with a good estimator.

To evaluate an estimator's overall performance, we look at its **average loss**, known as the **risk function**.

Risk Function $R(\theta, \delta)$

The risk of an estimator δ is its expected loss, where the expectation is taken over the distribution of the data X given θ.

$$R(\theta, \delta) = \mathbb{E}_\theta[L(\theta, \delta(X))]$$

Example: Risk of the Sample Mean

Let's find the risk for our coin toss example.

- Data: $X \sim \text{Binom}(100, \theta)$
- Estimator: $\delta(X) = \frac{X}{100}$
- Loss Function: Squared error loss $L(\theta, d) = (\theta - d)^2$

Example: Risk of the Sample Mean

Let's find the risk for our coin toss example.

- Data: $X \sim \text{Binom}(100, \theta)$
- Estimator: $\delta(X) = \frac{X}{100}$
- Loss Function: Squared error loss $L(\theta, d) = (\theta - d)^2$

The risk function $R(\theta, \delta)$ is:

$$\begin{aligned} R(\theta, \delta) &= \mathbb{E}_\theta\left[\left(\theta - \frac{X}{100}\right)^2\right] \\ &= \frac{1}{100^2}\mathbb{E}_\theta[(X - 100\theta)^2] \\ &= \frac{1}{100^2}\text{Var}_\theta(X) \\ &= \frac{100\theta(1-\theta)}{100^2} = \frac{\theta(1-\theta)}{100} \end{aligned}$$

A Fundamental Problem in Estimation

How do we choose the "best" estimator?

A fundamental problem arises when comparing two different estimators, say δ_1 and δ_2, using their risk functions.

The Challenge

If the risk functions $R(\theta, \delta_1)$ and $R(\theta, \delta_2)$ cross, there is no clear answer as to which estimator is universally better. One may have lower risk for some values of θ, while the other is better for other values of θ.