Families of Distributions

In statistics we are interested in some families of distributions, i.e., some collections of distributions.

For example, the family of binomial distributions with $p \in (0,1)$ and a fixed n; the family of normal distributions with $\mu \in \mathscr{R}$ and $\sigma > 0$.

Exponential families

A family of pdfs or pmfs indexed by θ is called an exponential family iff it can be expressed as

$$f_\theta(x) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right), \quad \theta \in \Theta,$$

where $\exp(x) = e^x$, Θ is the set of all values of θ (parameter space), $h(x) \geq 0$ and $t_1(x),...,t_k(x)$ are functions of x (not depending on θ), and $c(\theta) > 0$ and $w_1(\theta),...,w_k(\theta)$ are functions of the possibly vector-valued θ (not depending on x).

Note that the expression for f may not be unique.

Example 3.4.1.

To show that a family of pdf's or pmf's is an exponential family, we must identify the functions $h(x)$, $t_i(x)$, $c(\theta)$, and $w_i(\theta)$ and show that the pdf or pmf has the given form.

The *binomial*(n,p) distribution with $p \in (0,1)$ and a fixed n has pmf

$$\binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} (1-p)^n \exp\left(\log\left(\frac{p}{1-p}\right) x\right), \quad x = 0,1,...,n.$$

Let $\theta = p$, $c(\theta) = (1-p)^n$, $w_1(\theta) = \log(\frac{p}{1-p})$, $t_1(x) = x$, and $h(x) = \binom{n}{x}$ for $x = 0,1,...,n$ and $= 0$ otherwise.

Then, the binomial family with $p \in (0,1)$ and a fixed n is an exponential family $(k = 1)$.
(Note that $p = 0$ and $p = 1$ are not included in the family.)

Other examples: Poisson, negative binomial, normal, gamma, beta,...

Exponential families have many nice properties.
The following result is useful since we can replace integration or summation by differentiation.

Theorem 3.4.2.

If X has a pdf or pmf from an exponential family and $w_i(\theta)$'s are differentiable functions, then

$$E\left(\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X)\right) = -\frac{\partial \log c(\theta)}{\partial \theta_j}$$

where θ_j is the jth component of θ, and

$$\mathrm{Var}\left(\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X)\right) = -\frac{\partial^2 \log c(\theta)}{\partial \theta_j^2} - E\left(\sum_{i=1}^{k} \frac{\partial^2 w_i(\theta)}{\partial \theta_j^2} t_i(X)\right)$$

Proof.

From the exponential family expression for $f_\theta(x)$,

$$\log f_\theta(X) = \log h(X) + \log c(\theta) + \sum_{i=1}^{k} w_i(\theta) t_i(X)$$

Differentiating this expression leads to

$$\frac{\partial \log f_\theta(X)}{\partial \theta_j} = \frac{\partial \log c(\theta)}{\partial \theta_j} + \sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X)$$

Taking expectation, we obtain

$$E\left(\frac{\partial \log f_\theta(X)}{\partial \theta_j}\right) = \frac{\partial \log c(\theta)}{\partial \theta_j} + E\left(\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X)\right)$$

If $f_\theta(x)$ is a pdf (the proof for pmf is similar), then the left side of the previous expression is

$$\int_{-\infty}^{\infty} \frac{\partial \log f_\theta(x)}{\partial \theta_j} f_\theta(x)dx = \int_{-\infty}^{\infty} \frac{\partial f_\theta(x)}{\partial \theta_j} dx = \frac{\partial}{\partial \theta_j} \int_{-\infty}^{\infty} f_\theta(x)dx = \frac{\partial 1}{\partial \theta_j} = 0$$

We interchanged the differentiation and integration, which is justified under the exponential family assumption.
This proves the first result.

Note that

$$\frac{\partial^2 \log f_\theta(X)}{\partial \theta_j^2} = \frac{\partial}{\partial \theta_j}\left[\frac{\frac{\partial f_\theta(X)}{\partial \theta_j}}{f_\theta(X)}\right] = \frac{\frac{\partial^2 f_\theta(X)}{\partial \theta_j^2}}{f_\theta(X)} - \left[\frac{\frac{\partial f_\theta(X)}{\partial \theta_j}}{f_\theta(X)}\right]^2$$

Then

$$
\begin{aligned}
E\left(\frac{\partial^2 \log f_\theta(X)}{\partial \theta_j^2}\right) &= \int_{-\infty}^{\infty} \left\{ \frac{\frac{\partial^2 f_\theta(X)}{\partial \theta_j^2}}{f_\theta(X)} - \left[\frac{\frac{\partial f_\theta(X)}{\partial \theta_j}}{f_\theta(X)}\right]^2 f_\theta(x) \right\} dx \\
&= \int_{-\infty}^{\infty} \frac{\partial^2 f_\theta(X)}{\partial \theta_j^2} dx - \int_{-\infty}^{\infty} \left[\frac{\partial \log f_\theta(X)}{\partial \theta_j}\right]^2 f_\theta(x) dx \\
&= -\int_{-\infty}^{\infty} \left[\frac{\partial \log c(\theta)}{\partial \theta_j} + \sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X)\right]^2 f_\theta(x) dx \\
&= -\mathrm{Var}\left(\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta_j} t_i(X)\right)
\end{aligned}
$$

which follows from the first result.

Then the second result follows from

$$
\frac{\partial^2 \log f_\theta(X)}{\partial \theta_j^2} = \frac{\partial^2 \log c(\theta)}{\partial \theta_j^2} + \sum_{i=1}^{k} \frac{\partial^2 w_i(\theta)}{\partial \theta_j^2} t_i(X)
$$

Example 3.4.4.

If $X \sim N(\mu, \sigma^2)$, then $\theta = (\mu, \sigma)$ and

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(\frac{\mu x}{\sigma^2} - \frac{x^2}{2\sigma^2}\right)$$

Let $h(x) = 1$, $c(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{\mu^2}{2\sigma^2})$, $w_1(\theta) = 1/\sigma^2$, $w_2(\theta) = \mu/\sigma^2$, $t_1(x) = -x^2/2$, and $t_2(x) = x$.

Then this normal family is an exponential family with $k = 2$.

Applying Theorem 3.4.2, we obtain $E(X) = \mu$ from equation

$$-\frac{\partial \log c(\theta)}{\partial \mu} = \frac{\mu}{\sigma^2} = E\left(\sum_{i=1}^{2} \frac{\partial w_i(\theta)}{\partial \mu} t_i(X)\right) = E\left(\frac{X}{\sigma^2}\right)$$

Also,

$$-\frac{\partial \log c(\theta)}{\partial \sigma} = \frac{\mu^2}{\sigma^3} + \frac{1}{\sigma} = E\left(\sum_{i=1}^{2} \frac{\partial w_i(\theta)}{\partial \sigma} t_i(X)\right) = E\left(\frac{X^2}{\sigma^3} - \frac{2\mu X}{\sigma^3}\right)$$

Using $E(X) = \mu$, we obtain from this equation that $\mathrm{Var}(X) = \sigma^2$.

Beta distribution $beta(\alpha, \beta)$

For constants $\alpha > 0$ and $\beta > 0$, the *beta*(α, β) distribution has pdf

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

This is a pdf because

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

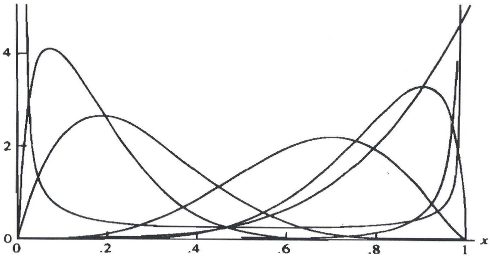

Figure 3.3.3. *Beta densities*

- Since $\Gamma(2) = \Gamma(1) = 1$, *beta*$(1,1)$ is the same as *uniform*$(0,1)$.
- If $X \sim$ *beta*(α,β), then $1 - X \sim$ *beta*(β,α).
- The pdf of *beta*(α,β) can be increasing ($\alpha > 1, \beta = 1$), decreasing ($\alpha = 1, \beta > 1$), U-shaped ($\alpha < 1, \beta < 1$), or unimodal ($\alpha > 1, \beta > 1$).
- If $\alpha = \beta$, then the pdf of *beta*(α,β) is symmetric about $\frac{1}{2}$.
- For any $r > 0$, if $X \sim$ *beta*(α,β), then

$$E(X^r) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{r+\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha+\beta)\Gamma(r+\alpha)}{\Gamma(\alpha)\Gamma(r+\alpha+\beta)}$$

  In particular,

$$E(X) = \frac{\alpha}{\alpha+\beta}, \qquad E(X^2) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

  Then

$$\mathrm{Var}(X) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

The family of *beta*(α,β) distributions is an exponential family.

Natural exponential family

If $\eta_i = w_i(\theta)$, $i = 1, ..., k$, and $\eta = (\eta_1, ..., \eta_k)$, the form of f_θ in the exponential family becomes

$$f^*_\eta(x) = h(x)c^*(\eta)\exp\left(\sum_{i=1}^{k} \eta_i t_i(x)\right)$$

η is called the natural parameter.
The set of η's for which $f^*_\eta(x)$ is a well-defined pdf is called the natural parameter space.

Full or curved exponential families

In an exponential family, if the dimension of θ is k (there is an open set $\subset \Theta$), then the family is a full exponential family. Otherwise the family is a curved exponential family.

An example of a full exponential family is $N(\mu, \sigma^2)$, $\mu \in \mathscr{R}$, $\sigma > 0$.
An example of a curved exponential family is $N(\mu, \mu^2)$, $\mu \in \mathscr{R}$.

How to show a family is not an exponential family

It may be difficult to show that a family is not an exponential family.

We cannot say "we are not able to express $f_\theta(x)$ in the form of an exponential family".

If f_θ, $\theta \in \Theta$ is an exponential family, then

$$\{x : f_\theta(x) > 0\} = \{x : h(x) > 0\}$$

which does not depend on θ values.

This fact can be used to show a family is non-exponential, i.e., if $\{x : f_\theta(x) > 0\}$ depends on θ, then f_θ, $\theta \in \Theta$, is not an exponential family.

Consider the family of two parameters exponential distributions with pdf's

$$f_\theta(x) = \begin{cases} \lambda^{-1} e^{-(x-\mu)/\lambda} & x > \mu \\ 0 & x \le \mu \end{cases} \qquad \mu \in \mathscr{R},\ \lambda > 0$$

It is not an exponential family because

$$\{x : f_\theta(x) > 0\} = \{x : x > \mu\}$$

Definition 3.5.2 (location family)

Let $f(x)$ be a given pdf. The family of pdf's, $f(x-\mu)$, $\mu \in \mathscr{R}$, is called a location family with location parameter μ.

- Examples of location families are normal and Cauchy with location parameter $\mu \in \mathscr{R}$ and the other parameter σ fixed.
  Other examples are given later.
- The pdf $f(x-\mu)$ is obtained by shifting the entire curve $f(x)$ by an amount μ (see the figure) without changing the structure of $f(x)$.
- It can be shown that $X \sim f(x-\mu)$ iff $X = Z + \mu$ with $Z \sim f(x)$.

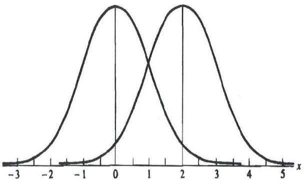

Figure 3.5.1. *Two members of the same location family: means at 0 and 2*

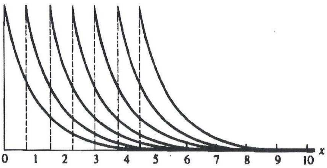

Figure 3.5.2. *Exponential location densities*

Definition 3.5.4 (scale family)

Let $f(x)$ be a given pdf. The family of pdf's, $\sigma^{-1} f(x/\sigma)$, $\sigma > 0$, is called a scale family with scale parameter σ.

- Examples of scale families are normal and Cauchy with scale parameter $\sigma > 0$ and μ fixed, *gamma*(α, β) with $\beta > 0$ and α fixed. Other examples are given later.
- The pdf $\sigma^{-1} f(x/\sigma)$ is obtained by stretching ($\sigma > 1$) or contracting ($\sigma < 1$) the curve $f(x)$ while still maintaining the same shape.
- It can be shown that $X \sim \sigma^{-1} f(x/\sigma)$ iff $X = \sigma Z$ with $Z \sim f(x)$.

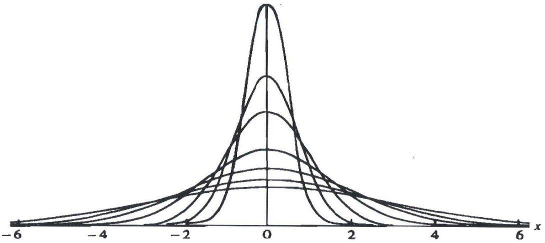

Figure 3.5.3. *Members of the same scale family*

Definition 3.5.5 (location-scale family)

Let $f(x)$ be a given pdf. The family of pdf's, $\sigma^{-1}f((x-\mu)/\sigma)$, $\mu \in \mathscr{R}$, $\sigma > 0$, is called a location-scale family with location parameter μ and scale parameter σ.

- A location-scale family is a combination of a location family and a scale family: it contains a sub-family that is a location family with any fixed σ, and a sub-family that is a scale family with any fixed μ.

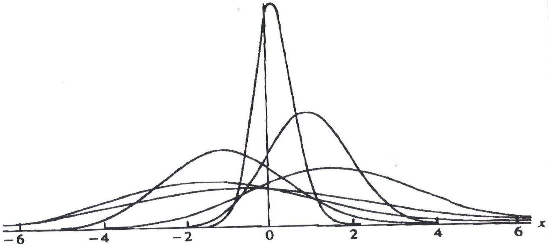

Figure 3.5.4. *Members of the same location–scale family*

- Examples of location-scale families are normal, double exponential, Cauchy, logistic, and two-parameter exponential distributions with location parameter $\mu \in \mathscr{R}$ and scale parameter $\sigma > 0$. Except for the two-parameter exponential distribution, all others are symmetric about μ.
- If $f(x)$ is symmetric about 0, then $\sigma^{-1}f((x-\mu)/\sigma)$ is symmetric about μ and μ is the median of $X \sim \sigma^{-1}f((x-\mu)/\sigma)$; furthermore, if the expectation of $f(x)$ exists, then μ is the expectation of $\sigma^{-1}f((x-\mu)/\sigma)$.
- It can be shown that $X \sim \sigma^{-1}f((x-\mu)/\sigma)$ iff $X = \sigma Z + \mu$ with $Z \sim f(x)$; furthermore, if $E(Z^2) < \infty$, then $E(X) = \sigma E(Z) + \mu$ and $\mathrm{Var}(X) = \sigma^2 \mathrm{Var}(Z)$.
- The pdf $f(x)$ in a location-scale family is standard iff the expectation $\int_{-\infty}^{\infty} xf(x)dx = 0$ and the variance $\int_{-\infty}^{\infty} x^2 f(x)dx = 1$.
- Typically, we choose a standard $f(x)$ to generate a location-scale family, in which case μ and σ^2 are the expectation and variance of $\sigma^{-1}f((x-\mu)/\sigma)$, respectively.

Two parameter exponential distribution *exponential*(μ,β)

If $X \sim$ *exponential*(β) and $\mu \in \mathscr{R}$ is a constant, then the distribution of $Y = X + \mu$ is called the two parameter exponential distribution and denoted by *exponential*(μ,β).
Its pdf and cdf are (by transformation)

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-(x-\mu)/\beta} & x \ge \mu \\ 0 & x < \mu \end{cases} \qquad F(x) = \begin{cases} 1 - e^{-(x-\mu)/\beta} & x \ge \mu \\ 0 & x < \mu \end{cases}$$

and, if $Y \sim$ *exponential*(μ,β),

$$E(Y) = \mu + \beta, \ \mathrm{Var}(Y) = \beta^2, \ M_Y(t) = \frac{e^{\mu t}}{1-\beta t}, \ t < \frac{1}{\beta}, \ \phi_Y(t) = \frac{e^{i\mu t}}{1 - i\beta t}, \ t \in \mathscr{R}$$

Double exponential distribution *double-exponential*(μ,σ)

By reflecting the pdf of *exponential*(μ,σ) around μ, we obtain the *double-exponential*(μ,σ) pdf that is symmetric about μ:

$$f(x) = \begin{cases} \frac{1}{2\sigma} e^{-(x-\mu)/\sigma} & x \ge \mu \\ \frac{1}{2\sigma} e^{(x-\mu)/\sigma} & x < \mu \end{cases} = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma}, \quad x \in \mathscr{R}$$

- This pdf is not bell-shaped; in fact, it has a peak (a non-differentiable point) at $x = \mu$.
- Its cdf is

$$F(x) = \begin{cases} 1 - \frac{1}{2} e^{-(x-\mu)/\sigma} & x \geq \mu \\ \frac{1}{2} e^{(x-\mu)/\sigma} & x < \mu \end{cases}$$

- If $X \sim \textit{double-exponential}(\mu, \sigma)$, then $Z = (X - \mu)/\sigma \sim \textit{double-exponential}(0, 1)$.
- If $Z = (X - \mu)/\sigma \sim \textit{double-exponential}(0, 1)$, then

$$E(Z) = \frac{1}{2} \int_{-\infty}^{\infty} x e^{-|x|} dx = 0$$

because $xe^{-|x|}$ is an odd function, and

$$\mathrm{Var}(Z) = E(Z^2) = \frac{1}{2} \int_{-\infty}^{\infty} x^2 e^{-|x|} dx = \int_0^{\infty} x^2 e^{-x} dx = \Gamma(3) = 2$$

- If $X \sim \textit{double-exponential}(\mu, \sigma)$, then $X = \sigma Z + \mu$, $Z = (X - \mu)/\sigma \sim \textit{double-exponential}(0, 1)$, and

$$E(X) = E(\sigma Z + \mu) = \mu, \qquad \mathrm{Var}(X) = \mathrm{Var}(\sigma Z + \mu) = \sigma^2 \mathrm{Var}(Z) = 2\sigma^2$$

Logistic distribution *logistic*(μ,σ)

For constants $\mu \in \mathscr{R}$ and $\sigma > 0$, the *logistic*(μ,σ) distribution has pdf

$$f(x) = \frac{e^{-(x-\mu)/\sigma}}{\sigma[1+e^{-(x-\mu)/\sigma}]^2}, \quad x \in \mathscr{R}$$

- This pdf is again bell-shaped and symmetric about μ.
- The cdf of *logistic*(μ,σ) has a close form:

$$F(x) = \int_{-\infty}^{x} f(t)dt = \frac{1}{1+e^{-(x-\mu)/\sigma}}, \quad x \in \mathscr{R}$$

- By symmetry, $E(X) = \mu$ if $X \sim$ *logistic*(μ,σ).
- The variance of $X \sim$ *logistic*(μ,σ) is not easy to obtain, but we give the result here: $\mathrm{Var}(X) = \sigma^2\pi^2/3$.

Pareto distribution *pareto*(α,β)

For constants $\alpha > 0$ and $\beta > 0$, the *pareto*(α,σ) distribution has pdf

$$f(x) = \begin{cases} \alpha\beta^\alpha x^{-(\alpha+1)} & x > \beta \\ 0 & x \le \beta \end{cases}$$

- First, f is indeed a pdf, because
$$\int_{-\infty}^{\infty} f(x)dx = \alpha\beta^{\alpha}\int_{\beta}^{\infty} x^{-(\alpha+1)}dx = \beta^{\alpha}x^{-\alpha}\Big|_{\infty}^{\beta} = \beta^{\alpha}\beta^{-\alpha} = 1$$
- Using a similar argument, we can obtain the cdf of $pareto(\alpha,\beta)$ as
$$F(x) = \begin{cases} 1-\left(\frac{\beta}{x}\right)^{\alpha} & x > \beta \\ 0 & x \le \beta \end{cases}$$
- Since the integral $\int_{\beta}^{\infty} x^{-t}dx$ is finite iff $t > 1$, $E(X) = \infty$ if $\alpha \le 1$ when $X \sim pareto(\alpha,\beta)$; if $\alpha > 1$, then
$$E(X) = \alpha\beta^{\alpha}\int_{\beta}^{\infty} x^{-\alpha}dx = \frac{\alpha\beta^{\alpha}}{\alpha-1}x^{-(\alpha-1)}\Big|_{\infty}^{\beta} = \frac{\alpha\beta^{\alpha}}{\alpha-1}\beta^{-(\alpha-1)} = \frac{\alpha\beta}{\alpha-1}$$
- Similarly, $\mathrm{Var}(X) = \infty$ if $\alpha \le 2$; and if $\alpha > 2$,
$$E(X^2) = \alpha\beta^{\alpha}\int_{\beta}^{\infty} x^{-\alpha+1}dx = \frac{\alpha\beta^{\alpha}}{\alpha-2}x^{-\alpha+2}\Big|_{\infty}^{\beta} = \frac{\alpha\beta^{\alpha}}{\alpha-2}\beta^{-\alpha+2} = \frac{\alpha\beta^2}{\alpha-2}$$
$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{\alpha\beta^2}{\alpha-2} - \frac{\alpha^2\beta^2}{(\alpha-1)^2} = \frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)}$$

Weibull distribution $Weibull(\gamma, \beta)$

For constants $\gamma > 0$ and $\beta > 0$, if $X \sim exponential(\beta)$, then $Y = X^{1/\gamma} \sim Weibull(\gamma, \beta)$ with pdf

$$f(x) = \begin{cases} \frac{\gamma}{\beta} x^{\gamma-1} e^{-x^\gamma/\beta} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

- An example of $Y \sim Weibull(\gamma, \beta)$ is lifetime or failure time.
- If $Y \sim Weibull(\gamma, \beta)$, then $X = Y^\gamma \sim exponential(\beta)$ and

$$\begin{aligned} E(Y) &= E(X^{1/\gamma}) = \frac{1}{\beta} \int_0^\infty x^{1/\gamma} e^{-x/\beta} dx \\ &= \beta^{1/\gamma} \int_0^\infty u^{1/\gamma} e^u du = \beta^{1/\gamma} \Gamma\left(\frac{1}{\gamma} + 1\right) \end{aligned}$$

Similarly, we can obtain that

$$\mathrm{Var}(Y) = \beta^{2/\gamma} \left\{ \Gamma\left(\frac{2}{\gamma} + 1\right) - \left[\Gamma\left(\frac{1}{\gamma} + 1\right)\right]^2 \right\}$$