پارامترها به‌وجود خواهد آورد. با حل این دستگاه، پارامترها برحسب گشتاورهای نمونه به‌دست می‌آیند.

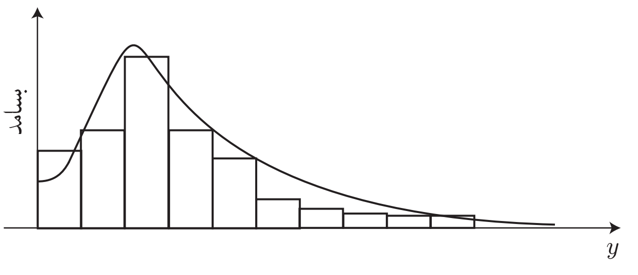
به طور کلی، همان‌طورکه در فصل ۳ دیدیم، گشتاور مرتبهٔ kام به مرکز صفر را در نمونه با $m'_k = \sum_{i=1}^{n} y_i^k/n$ نشان می‌دهیم. اگر نمونهٔ مورد استفاده از جامعه‌ای گرفته شده باشد که دارای r پارامتر مجهول است و گشتاور مرتبهٔ kام به مبدأ صفر از این جامعه را با $\mu'_k = E(Y^k)$ بنمایانیم، روش برآورد گشتاوری مستلزم حل دستگاه r معادلهٔ r مجهولی زیر است.

$$\mu'_k = m'_k, \qquad k = 1, \ldots, r \tag{1}$$

در این دستگاه m'_kها مقدارهای معلوم و μ'_kها مقدارهای مجهول معادله‌ها هستند.

مثال ۱

داده‌هایی که از آزمایش طول عمر خستگی ۷۵ نمونه از آلومینیم $S-T$ به‌دست آمده‌اند، دارای بافت‌نگار رسم شده در شکل زیر هستند. این بافت‌نگار حاکی از آن است که توزیع لگ نرمال $Y \sim \mathrm{Log}N(\nu, \tau^2)$ خوب به آن می‌پردازد. داده‌های نمونه منجر به $\overline{y}_{75} = 26{,}75$ (میلیون چرخه) و $s^2_{75} = 360{,}0$ شدند.



پارامترهای (τ, ν) و نیز پارامترهای (σ^2, μ) برای توزیع نرمال متناظر را به روش گشتاوری بیابید.

حل. در رابطهٔ (۲۹) بخش ۶-۲ دیدیم که اگر $Y \sim \log N(\nu, \tau^2)$ باشد، پارامترهای (σ^2, μ)ی توزیع نرمال متناظر عبارت‌اند از $\nu = E(Y) = \exp\left\{\mu + \frac{1}{2}\sigma^2\right\}$ و $\mu'_2 = E(Y^2) = \exp\left\{2(\mu + \sigma^2)\right\}$. پس

$$\begin{cases} \overline{y}_{75} = 26{,}75 = e^{\mu + \frac{1}{2}\sigma^2} \\ \frac{1}{75}\sum_{i=1}^{75} y_i^2 = e^{2(\mu+\sigma^2)} \end{cases}$$

جواب این دستگاه عبارت است از

$$\begin{cases} \widetilde{\mu} = ۳/۰۹ \\ \widetilde{\sigma}^۲ = ۰/۴۰ \end{cases}$$

از سوی دیگر، $\overline{y}_n = \widetilde{\nu}$ و $\widetilde{\tau}^۲ = \widetilde{\nu}^۲(e^{\widetilde{\sigma}^۲} - ۱) = (۲۶/۷۵)^۲(e^{۰/۴} - ۱)$، پس $\widetilde{\nu} = ۲۶/۷۵$ و $\widetilde{\tau}^۲ = ۳۵۱/۹۳$ یا $\widetilde{\tau} = ۱۸/۸$. □

در پیوست‌های ۱ و ۲ فصل ۸ برآوردگرهای توزیع‌های مهم را به‌دست آورده‌ایم.

روش برآورد گشتاوری با آنکه روشی ساده و ملموس است دو نقص مهم دارد: برابری گشتاورهای نمونه و جامعه در نمونه‌های بزرگ تا حدی درست است و در نمونه‌های کوچک این برابری دور از واقعیت است. نقص دوم آن است که این روش حافظ دامنهٔ پارامتر نیست. یعنی گاهی برآوردی را به‌دست می‌دهد که خارج از مجموعهٔ مقادیر ممکن برای پارامتر است. مثلاً، گاهی واریانس را منفی برآورد می‌کند. از این رو روشی چندان مطلوب نیست، اما در مواردی نادر تنها راه چاره است. یک مشکل دیگر درکاربست روش گشتاوری آن است که در برخی توزیع‌ها گشتاورهای جامعه نامتناهی‌اند.

۲-۸ روش ماکسیمم درستنمایی[1]

در روش ماکسیمم درستنمایی فرض بر این است که جامعهٔ مادر با تابع احتمال گسسته یا پیوستهٔ $f_Y(y|\theta)$ توصیف می‌شود که در آن نماد θ را برای معرفی پارامتر یا پارامترها به‌کار برده‌ایم. مثلاً در توزیع دوجمله‌ای داریم $p_Y(y|\pi)$، در نرمال $f_Y(y|\mu, \sigma^۲)$ و در پواسون $f_Y(y|\lambda)$ که همهٔ این صورت‌ها را با صورت کلی $f_Y(y|\theta)$ بیان می‌کنیم. وقتی یک نمونهٔ تصادفی سادهٔ nعضوی از این جامعه می‌گیریم، هر عضو نمونه بنابر مفروضات عمل نمونه‌گیری می‌تواند هر یک از عضوهای جامعه باشد که توزیع احتمال آن همانند توزیع جامعهٔ مادر و برابر با $f_Y(y|\theta)$ است. بنابر فرض استقلال عضوهای نمونه، احتمال مشاهدهٔ مقدارهای عددی خاص مانند $y_۱$، $y_۲$، ...، y_n در نمونه در توزیع‌های گسسته برابر با حاصل‌ضرب احتمال مشاهدهٔ یکایک آنهاست. اما در توزیع‌های پیوسته، چنانکه در تعریف چگالی احتمال دیدیم، احتمال افتادن عضو iام نمونهٔ (Y_i) در بازهٔ بینهایت کوچکی به مرکز y_i مانند $[y_i - dy_i/۲,\ y_i + dy_۱/۲]$ خیلی نزدیک به $f(y_i|\theta)dy_i$ است. پس احتمال توأم مشاهدهٔ مقدارهای عددی خاص بالا در نمونه، متناسب با حاصل‌ضرب $\pi_{i=۱}^n f(y_i|\theta)dy_i$ است. چون مقدار $k = \pi_{i=۱}^n dy_i$ به θ بستگی ندارد و بحث ما دربارهٔ θ است، می‌توانیم از k در بحث‌های زیر صرف‌نظر

1. Method of Maximum Likelihood (ML)

کنیم. اگر برای راحتی در نوشتن فرمول‌ها قرار بگذاریم $\mathbf{Y} = [Y_۱, \ldots, Y_n]$ و $\mathbf{y} = [y_۱, \ldots, y_n]$ باشد، و مقدارهای ممکن θ را که فضای پارامتر[1] می‌نامند با Θ نشان دهیم، تابع احتمال (تابع چگالی احتمال) توأم اعضای نمونه عبارت خواهد بود از

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}|\theta) &= f_{Y_۱}(y_۱|\theta) f_{Y_۲}(y_۲|\theta) \ldots f_{Y_n}(y_n|\theta) \\ &= \left[\prod_{i=۱}^{n} f_Y(y_i|\theta)\right] I_\Theta\ (\theta) \end{aligned} \qquad (۱)$$

که در آن θ مقداری از فضای Θ را اختیار می‌کند و مجهول است.[2]

رابطهٔ (۱) در واقع متناسب با این احتمال است که در نتیجهٔ نمونه‌گیری تصادفی ساده مقدارهای $y_۱$، $y_۲$، …، y_n در نمونه مشاهده شوند. در توزیع‌های گسسته ضریب تناسب $k = ۱$ و در توزیع‌های پیوسته $k = \prod_{i=۱}^{n} dy_i$ است. مثلاً در توزیع پواسون با نرخ رخداد مجهول λ، احتمال آنکه نمونهٔ $\mathbf{y} = [۳, ۵]$ مشاهده شود برابر است با $p = P[Y = ۳|\lambda] P[Y = ۵|\lambda]$ که اگر آن را به‌صورت (۱) بنویسیم، خواهیم داشت

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}|\lambda) &= \left[e^{-\lambda}\lambda^{۳}/۳!\right]\left[e^{-\lambda}\lambda^{۵}/۵!\right] \\ &= \left[e^{-۲\lambda}\lambda^{۸}/۳!۵!\right] I_{(۰,\infty)}(\lambda) \end{aligned} \qquad (۲)$$

یا در توزیع دوجمله‌ای با احتمال «پیروزی» مجهول π، اگر از بین $n = ۵$ نمونه $Y = ۲$ «پیروزی» حاصل شده باشد، داریم $P[Y = ۲|\pi] = \binom{۵}{۲}\pi^{۲}(۱-\pi)^{۳}$ که پیشاپیش به‌صورت (۱) است. با توجه به فضای پارامتر $۱ > \pi > ۰$، احتمال اخیر که به π بستگی دارد عبارت است از

$$f_Y(y|\pi) = \left[\binom{۵}{۲}\pi^{۲}(۱-\pi)^{۳}\right] I_{(۰,۱)}(\pi) \qquad (۳)$$

در (۱) یا یکی از مصداق‌های ویژهٔ آن توزیع دوجمله‌ای (۳)، پارامتر π مجهول است و هدف نمونه‌گیری برآورد آن از روی مقادیر مشاهده شدهٔ نمونه $\mathbf{y}$ است. ملاحظه می‌کنیم که احتمال مشاهدهٔ $\mathbf{y}$ بستگی به مقدار π دارد. اینک ممکن است از خود بپرسیم که چه مقداری از π بیشترین احتمال را برای مقدار عملاً

1. Parameter Space

۲. در موارد نادر، فضای پارامتر به $\mathbf{y}$ بستگی دارد که باید به‌صورت $\Theta_{\mathbf{y}}$ نوشت، اما برای سادگی نمادگذاری اندیس y را نخواهیم نوشت.

مشاهده شدهٔ y فراهم می‌سازد. طبیعی است انتظار داشته باشیم که از بین مقادیر ممکن y آن مقدار که ماکسیمم احتمال را دارد مشاهده شود. پس با ماکسیمم کردن (۳) برحسب θ می‌توان برآوردی معقول از θ را به‌دست آورد. این روش را روش ماکسیمم درستنمایی و برآورد حاصل را برآورد ماکسیمم درستنمایی[1] گویند. صورت تابعی برآورد را برآوردگر ماکسیمم درستنمایی می‌نامند. بنابراین، به‌دست آوردن برآورد ماکسیمم درستنمایی θ که آن را با $\hat{\theta}$ نشان می‌دهیم، مستلزم ماکسیمم کردن تابع (۱) برحسب θ است. توجه کنید که در (۱) تنها متغیر در تابع همان پارامتر θ است زیرا y مقداری است که عملاً در نمونه مشاهده شده و دیگر نمی‌تواند تغییر کند، بلکه ثابت است. برای تأکید بر این نکته که $f_{\mathbf{Y}}(\mathbf{y}|\theta)$ به‌ازای مقدار مشاهده شدهٔ $\mathbf{y}$ تابع θ است آن را با نماد $L(\mathbf{y}|\theta) = f_{\boldsymbol{Y}}(\boldsymbol{y}|\theta)$ نشان می‌دهند و آن را تابع درستنمایی[2] گویند.

پیش از بیان قاعدهٔ کلی روش برآورد ماکسیمم درستنمایی، آموزنده خواهد بود که حالت خاص (۲) را به‌دقت بررسی کنیم.

احتمال (۲) را به‌ازای چند مقدار از λ حساب می‌کنیم و در جدول ۱ می‌آوریم.

جدول ۱ چند مقدار از تابع درستنمایی توزیع پواسون با پارامتر λ

λ / $L(\mathbf{y}\|\lambda)$	۰	۱	۲	۳	۴	۵	۶
احتمال مشاهدهٔ نمونهٔ (۵, ۳)	$\frac{۰}{۷۲۰}$	$\frac{۰/۱۳۵}{۷۲۰}$	$\frac{۴/۶۸۹}{۷۲۰}$	$\frac{۱۶/۲۶۳}{۷۲۰}$	$\frac{۲۱/۹۸۵}{۷۲۰}$	$\frac{۱۷/۷۳۴}{۷۲۰}$	$\frac{۱۰/۳۲۰}{۷۲۰}$

مشاهده می‌شود که با این داده‌ها تابع درستنمایی $L(\mathbf{y}|\lambda) = e^{-۲\lambda}\lambda^{۸}/۷۲۰$ در نقطهٔ $\lambda = ۴$ ماکسیمم مقدار خود را دارد. پس بنابر قاعدهٔ درستنمایی که در بالا بیان شد، باید $\hat{\lambda} = ۴$ را برآورد ماکسیمم درستنمایی λ بگیریم. این نتیجه را به روش تحلیلی نیز می‌توان به‌دست آورد. در واقع، تابعی از λ داریم که می‌خواهیم ماکسیمم کنیم. پیداست که در این مثال $L(\mathbf{y}|\lambda)$ تابعی پیوسته از λ در بازهٔ $(۰, \infty)$ است. به‌ازای مقدارهای حدی $\lambda = ۰$ و $\lambda = \infty$ تابع صفر است. به‌ازای دیگر مقادیر λ، تابع مثبت است. پس باید ماکسیمم داشته باشد. نسبت به λ مشتق می‌گیریم:

$$L'(\mathbf{y}|\lambda) = \frac{۱}{۷۲۰}\left[-۲e^{-۲\lambda}\lambda^{۸} + ۸\lambda^{۷}e^{-۲\lambda}\right] = \frac{۲\lambda^{۷}e^{-۲\lambda}}{۷۲۰}[۴ - \lambda]$$

چون $\lambda > ۰$ است ریشهٔ $L'(\mathbf{y}|\lambda) = ۰$ برابر با $\hat{\lambda} = ۴$ است. برای آنکه دریابیم این مقدار طول

1. Maximum Likelihood Estimate (MLE)   2. Likelihood function